CHARACTERISATION OF HFC-134a BY GAS CHROMATOGRAPHY-MASS SPECTROMETRY AND CHEMOMETRICS

A thesis submitted in fulfilment of the requirements for the award of the degree

of

DOCTOR OF PHILOSOPHY

from

THE UNIVERSITY OF EDINBURGH

by

MICHAEL ANTHONY REILLY, B.Sc. (Hons.)

Department of Chemistry
1999

DECLARATION

I hereby certify that the work presented in this thesis is original, except where due acknowledgement has been made in the text of the thesis, and has not been submitted for a higher degree at any other university or institution

Dedicated to my family

ABSTRACT

1,1,1,2-tetrafluoroethane (HFC 134a) is one of the major replacement compounds for chlorofluorocarbons and is therefore an important industrial chemical. This work describes a method of characterisation for HFC 134a. The method profiled the levels of synthesis by-products present in samples of HFC 134a using gas chromatography (GC) with detection by flame ionisation detector (FID) or electron capture detector (ECD). The principal method for identification of the by-products was EI GC-mass spectrometry. The multi-variate data produced by the profiling of samples were analysed using chemometric techniques.

A training set of samples of HFC 134a, with known origins of production, was analysed by both GC-FID and GC-ECD. This training data set was used to investigate the different methods of chemometric analysis as applied to the raw data, normalised data and principal component analysed data. K-means clustering and Hierarchical clustering were investigated to find the optimum methods for the identification of samples' origins of production based on their chromatographic profiles.

The FID chromatographic traces could be correctly identified by applying a two step principal component analysis (PCA) using a hierarchical clustering method to classify the samples. The ECD chromatographic data could be correctly identified by applying a PCA followed by classification using a hierarchical clustering method.

These clustering techniques were then tested by classifying a second set of samples of HFC 134a with known production origins. The chromatographic traces for the FID and ECD analyses of the test samples were correctly identified. The two classification techniques were used to identify further samples into clusters belonging to the known origins of production or into new clusters representing samples of unknown origin. Classification of samples using the ECD data required the least amount of operator interpretation and provided the least amount of ambiguity in sample identification.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisors, colleagues and friends who have provided invaluable assistance in the course of this study;

- to my supervisor, Prof. John Monaghan, for constant help and good humour which helped me through the difficult times.
- to my second supervisor, Dr Mat Heal, for pushing me along and posing questions that I needed to find answers to.
- to my industrial supervisors, Dr David Barnes and Mr Alan Handley, for supplying me with guidance and much needed equipment
- to the EPSRC for providing the funding to carry out the research.
- to ICI C&P Runcorn for providing financial assistance in the course of this work.
- to the BMSS who provided the funding for me to present my work at a number of different scientific forums.
- to Alan and Brian who constantly furnished me with pieces of equipment and practical knowledge without which progress would have been impossible.
- to Kathryn, who has helped so much in the production of this work and kept me sane no matter how stressed I became.
- to Stavros and Little Lord Fontelroy, you know who you are and how many good times we have all had in Edinburgh, you will always be my friends.
- to Dr Andrew Penman who sent me down this dark path of analytical chemistry, can I ever forgive you!
- to my parents, who furnished me with the principals and ideals which have allowed me to progress successfully through my life.

TABLE OF CONTENTS

Glossary of terms

AC	Alternating electrical current.
Ashai Glass	HFC 134a sample produced by Ashai Glass.
Atochem	HFC 134a sample produced by Atochem.
by-product	Compound produced in synthetic pathway employed to form required compound.
CFC	Chlororfluorocarbon.
DC	Direct current.
e^-	Single electron.
ECD	Electron capture detector.
EI	Electron impact ionisation.
FID	Flame ionisation detector.
GC	Gas chromatography.
GC-MS	Gas chromatography mass spectrometry.
GWP	Global warming potential.
HETP	Height per theoretical plate.
HCFC	Hydrochlorofluorocarbon.
HFC	Hydroflurocarbon.
Hoechst	HFC 134a sample produced by Hoechst.
λV	Light in the ultra violet region.
IR	Light in the infra red region.
Japan	HFC 134a sample produced by ICI in Japan.
LOD	Limit of detection.
LOQ	Limit of quantification.
MS	Mass spectrometry.
m/z	Mass-to-charge ratio. Used for singly and multiply charged ions.
n	Number of observations.
ODP	Ozone depleting potential.
PC	Principal components.
PCA	Principal component analysis.

Rf Alternating current in the radio frequency.
RRt Relative retention time.
RSD Relative standard deviation
RT Retention time.

Showa Denka HFC 134a sample produced by Showa Denka.

TIC Total ion current. Analogous to the chromatogram in chromatography. Generated by summing the ion currents in each scan of the mass analyser.
TFA Trifluoroacetic acid.

UK HFC 134a sample produced by ICI in the UK.
USA HFC 134a sample produced by ICI in the USA.
UV Light in the ultra violet region.

Chapter 1

Introduction

1.0 Thesis overview

This thesis details the work performed to develop a method to characterise samples of 1,1,1,2-tetrafluoroethane (HFC 134a). The central question addressed was, is it possible, using standard analytical techniques, to identify and distinguish samples of HFC 134a on the basis of their routes of synthesis and origins of production. The samples are to be characterised by looking at the trace levels of synthesis by-products which are present in all samples of HFC 134a.

The analytical approach employed used gas chromatography (GC), to separate the by-products from HFC 134a and flame ionisation detection (FID) and/or an electron capture detection (ECD) to profile the by-products. Identification of the by-products was performed using GC-Mass Spectrometry. The multi-variate data produced in these analyses were investigated using chemometric techniques which can highlight patterns and structures in the data.

1.1 Thesis structure

The thesis is structured into eight chapters. Chapter 1 is principally an introduction to the history and development of HFC 134a and its role as a major commercially produced compound. A brief history of chloroflurorcarbons (CFC) and the environmental problems which led to the phasing out of CFCs and the development of HFC 134a as a replacement is given. This is followed by an overview of the analytical techniques to be employed, detailing parameters to be optimised and potential problem areas to be considered in the analysis of HFC 134a samples.

Chapter 2 gives a brief introduction to chemometrics and the multivariate statistics it involves. A short account of the development of chemometrics over recent years is given, followed by an introduction to the theory and application of the techniques used in this work. The account is meant only as a starting point in the field of chemometrics and is by no means a comprehensive review of the subject.

Chapter 3 is concerned with the experimental design and set-up employed to analyse samples of HFC 134. It gives a full account of all the experimental parameters which have to be monitored and optimised in order to produce the most accurate and precise data. These range from investigating sample dilution methods to effects of temperature on detector response.

Chapter 4 lists the results obtained for the analyses of the samples of HFC 134a by both ECD and FID. Also included are the identities and postulated structures of the by-products present elucidated using GC-MS.

Chapters 5 details the way in which the raw data was transformed in order to facilitate the chemometric analysis of the data produced by the chromatographic profiling of the samples of HFC 134a analysed by GC with detection by FID and ECD.

Chapter 6 describes how the transformed data was analysed by chemometrics. The chapter presents the classification results obtained by applying a number of different clustering techniques to the different transformed data. The chapter concludes by summing up which classification protocol provided the best classification of the test samples which were all of known origin.

The classification protocols which were shown to give the best results in Chapter 6 were then used to identify a further set of samples. Chapter 7 details the results of classifying these further samples of HFC 134a of known origin, mixed samples of known origin and samples of unknown origin.

Chapter 8 is a final discussion chapter which draws together the conclusions formed to show how the techniques employed have characterised the samples of HFC 134a analysed. Also included are ideas for continuing work to further test and improve the characterisation method and details of areas which require investigation as a consequence of this work.

1.2 History of the chlorofluorocarbons

Chlorofluorocarbons (CFCs) were developed in the 1920s by General Motors Research Laboratories as alternative non-toxic non-flammable refrigerants to replace sulfur dioxide and ammonia in use at the time[1] (see Appendix A for nomenclature scheme). The chemical inertness of CFCs[2] led to their widespread use as aerosol propellants, blowing agents for foam plastic production and solvents as well as their intended use in refrigeration and air conditioning units. CFC production quickly became a worldwide industry with an estimated usage, in 1986, in the USA alone of 318 million pounds sterling[3]. The uses of CFCs ultimately led to their release into the atmosphere *e.g.* aerosol propellants, leaks from hermetically sealed refrigeration units at the end of their working lives. However, this was not thought to be a problem as the CFCs were inert gases and had no known adverse effects on the environment[1].

Measurements by Lovelock in 1973 showed levels of CFCs in the atmosphere to be equal, to within experimental error, to the total amount thought to have been released into the environment[4]. The expected tropospheric inertness of the gases had been confirmed and lifetimes of the order of 10s to 100s of years were indicated. It was not until a year later when Rowland and Molina, while investigating the effect of space rocket emissions on the environment, highlighted a potential threat to stratospheric ozone that concerns were raised over CFCs[5].

The major breakdown pathway for the CFCs is through ultraviolet photolysis (Equation 1.1) at wavelengths of 190-230nm[5]. The breakdown mainly occurs at high atmospheric altitudes in the stratosphere as the majority of the ultraviolet light is absorbed above the stratopause[6].

$$CF_2Cl_2 + h\nu \rightarrow CF_2Cl + Cl \qquad \textbf{Equation 1.1}$$

The chlorine radical formed in this reaction can react with ozone (O_3) in a catalytic cycle (Equation 1.2). This reaction has the net effect of converting O and O_3 into O_2 and so lowers the concentration of O_3 present in the statosphere.

$$O_3 + Cl \rightarrow ClO + O_2$$
$$ClO + O \rightarrow Cl + O_2$$
$$O_3 + O \rightarrow O_2 + O_2$$

Equation 1.2

The mixing ratio of O_3 to other gases varies dramatically in the atmosphere depending on altitude (Figure 1.1). The rate of formation of O_3 depends on the level of ultraviolet light and concentration of O_2 in the atmosphere. The concentration of O_3 in the atmosphere follows a Chapman profile[7]. That is to say, in the upper stratosphere there are high levels of *hv* but low levels of O_2, consequently little O_3 is formed. In the lower stratosphere there are high levels of O_2 but low levels of *hv*, consequently little O_3 is formed. It is at mid levels in the stratosphere where the product of O_2 concentration and *hv* intensity are at maximum that the highest levels of O_3 are found. This forms what is known as the 'ozone layer'[8].

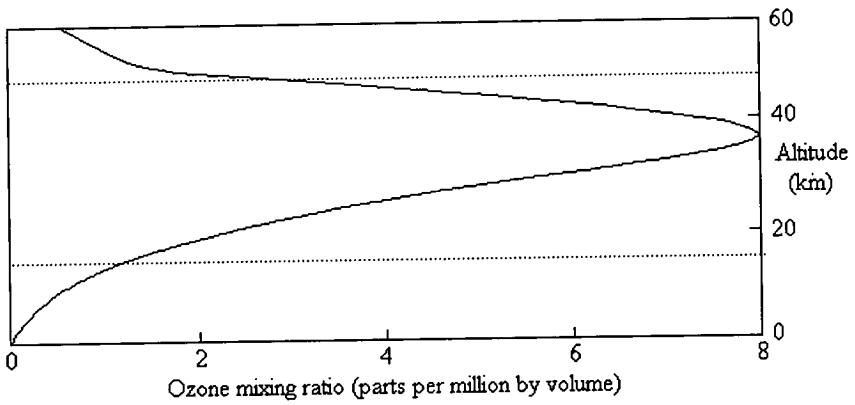


Figure 1.1: *Variation of atmospheric ozone concentration with altitude as a relative mixing ratio.*

The concentration of ozone in the atmosphere is small. However, ozone has an extremely important role in Earth's environmental climate. First, ozone absorbs virtually all solar ultraviolet light in the wavelength region 240 - 290nm[7] which would otherwise be transmitted to the Earth's surface. Radiation at these wavelengths is lethal to unicellular organisms and to the surface cells of plant and animal life[9]. Secondly, upper atmospheric meteorology is greatly influenced by the

heating of ozone that occurs after its absorption of ultraviolet (UV), infra red (IR) and visible radiation.

After the discovery of the potential risk of release of CFCs into the atmosphere and their subsequent effect on ozone concentration, steps were taken to limit and reduce their use and release. It was not until 1984, however, that the true impact of these compounds on atmospheric ozone was fully realised. Research measurements of ozone levels by the British Antarctic Survey showed dramatic losses of ozone over Antarctica[10]. The overall loss of ozone concentration was 30% over the period of 1977-84 and the rate of loss of ozone was increasing. Most dramatic of all, however, was a climatic phenomenon over Antarctica which led to a rapid and large loss of ozone (up to 95%) every spring forming what is now known as the 'ozone hole'.

Following the publication of this work and further studies confirming the results, a period of complacency over the ozone layer came to an end. The unexpected depletion levels of ozone led to far reaching proposals for the reduction and eventual phasing out of the use of CFCs drawn up in an international meeting on the environment in Montreal[11]. Further revisions and amendments led to the complete phasing out of all CFCs and some hydrochlorofluorocarbons (HCFCs) along with a number of other known ozone depleting compounds. Although the restrictions on production and use of these compounds are necessary, they do present the major problem of finding replacements. As discussed earlier, the uses of these compounds are widespread and the time scales set by governments for their phasing out meant alternatives had to be developed quickly.

1.3 HFC 134a as a CFC alternative

With the phasing out of use of CFCs industry was faced with a major challenge. In a relatively short period of time, approximately 10 years, replacements had to be found. They had to be non-toxic, non-flammable, compatible with existing technologies and to show none of the environmental problems of CFCs. For these reasons industry focused on developing and assessing hydrochlorofluorocarbons (HCFCs) and hydrofluorocarbons (HFCs)[12]. Although HCFCs have greatly reduced effects on the

environment compared with CFCs they are still considered only as a temporary solution and used only where no other alternative compound can be presently employed.

One of the potential longer-term replacement compounds for CFCs is the HFC, 1,1,1,2-tetraflurorethane (HFC 134a)[13]. HFC 134a is used as the main replacement compound for CFC 12, one of the most widely used CFCs (*e.g.* refrigeration and air conditioning units). The thermodynamic properties of HFC 134a are very similar to those of CFC 12. However, there are a number of other practical considerations to be taken into account, such as compatibility with lubricating oils within refrigeration units. HFC 134a was found not to have the same levels of miscibility with lubricating oils used in cooling units as CFC 12. This required alternative oils to be developed and used and this is still an area of active research. Furthermore, HFC 134a was found to be an extremely potent extracting solvent and had the undesirable effect of breaking down seals and hosing used to enclose the cooling units. For these reasons HFC 134a cannot be considered as a straight drop-in replacement for CFC 12.

As well as the practical considerations for the use of HFC 134a its potential effect on the environment had to be considered. The main areas to be investigated were its atmospheric lifetime, effect on the ozone layer, effect on global warming and any risk from its atmospheric breakdown products.

1.3.1 Fate of HFC 134a in the environment

The effect of HFC 134a on the ozone can be expressed as its ozone depletion potential (ODP)[14]. This is a ratio value calculated by comparing the effect of any specific halocarbon on ozone levels compared with that of CFC 11. In the case of HFC 134a, as the compound contains no chlorine atoms, it has an ODP of zero.

The effect of HFC 134a on global warming is likewise expressed as a ratio value compared with CFC 11. Although the IR absorption profile of HFC 134a is very

similar to that of CFC 11 the greatly reduced atmospheric lifetime of HFC 134a means that the global warming potential (GWP)[15] of HFC 134a is a number of orders of magnitude lower than that of CFC 11.

The shortened atmospheric lifetime of HFC 134a over the CFCs it replaces results from its reaction with hydroxyl (OH) radicals. The overall atmospheric breakdown pathway is shown in Scheme 1.1[16].

Initial breakdown of HFC 134a is mediated by attack by OH radicals removing hydrogen to form an alkyl radical. An alkyl peroxy radical is formed by combination with molecular oxygen on collision with a molecular body, which removes excess energy from the peroxy radical. Reaction with nitric oxide (NO), which is oxidised to nitrogen dioxide (NO_2), forms the alkoxy radical[17]. The alkoxy radical can then be further decomposed or react with molecular oxygen to form an acetyl halide. Hydrolysis of the acetyl halide after uptake into aqueous environmental media forms trifluoroacetic acid (TFA). The reaction yield of TFA from HFC 134a is in the region of 7 - 20%. The rate of reaction and overall yield of TFA are dependent on a number of environmental conditions such as hydroxyl radical and nitric oxide concentrations, which are elevated in urban areas.

Scheme 1.1

$$CF_3CH_2F \longrightarrow CF_3CHF \longrightarrow CF_3CHFOO \longrightarrow CF_3CHFO \longrightarrow FCF(O)F$$

The low Henry's Law constant for TFA[18] means that it will partition into the aqueous phases that occur throughout the environment. Rapid removal from the atmosphere by precipitation means that TFA levels will accumulate in ground based water supplies. TFA is resistant to abiotic degradation processes and is virtually unmetabolised by most plants and animals. Although experimental systems have shown that TFA is broken down by microbes under both oxic and anoxic conditions[19] the actual extent of removal from the environment by these processes is unknown. The potential slow breakdown of TFA in ground water supplies leads to areas of concern as to its effects on flora and fauna. Toxicological data for TFA are

limited but it is known to have detrimental effects on plant growth, especially on vascular plants where it can be concentrated up to 30 times in the xylem tissue[20].

Early modelling and experimental data showed that levels of TFA in ground water would reach only nominal concentration levels[21]. Even by the year 2010 (time for expected maximal use of HFC 134a) these concentration levels would still be orders of magnitude below those thought to be harmful to life. However, concern has been rising[22] as experimental measurements by a number of groups have shown that the levels of TFA in precipitation and ground water are much higher than would be expected[23]. Further to this, certain areas have been found to have higher than average levels of TFA[24]. Urban areas which have raised levels of HFC 134a release and higher levels of OH radicals, because of pollution, have TFA levels in precipitation orders of magnitude higher than those in the countryside. Also, certain wetlands have been shown to have higher levels of TFA because of reduced cycling of water or concentration effects (as seen with evaporation in the Dead Sea).

Further work is required in the areas of modelling and measuring the levels of HFC 134a and its breakdown products. Studies are continuing into the toxic effect of TFA on plant and animal life and the possible routes of removal of TFA from the environment. Obviously a transition from the use of CFCs is required and will be continued. However, close monitoring of the effects of HFC 134a and its break down products are required to prevent any major environmental problems.

1.3.2 Large scale production

As with any compound there are a number of different synthetic routes of production for HFC 134a. The key to changing a reaction in the laboratory into a full scale production run is the concept of scale up[25]. Careful consideration must be taken over every step in a possible industrial scale synthetic route as what may be acceptable in a laboratory reaction may not be feasible in a large scale plant. The large scale production method chosen by any company needs to fulfil a great many parameters. The method must minimise the steps taken to form the final product, limit

temperature and pressures, avoid using costly or dangerous raw materials, review type and amounts of waste products *etc.* The considerations are numerous and often conflicting in many circumstances. The companies producing HFC 134a whose samples have been considered in this work use one of two general methods[26].

1.3.3 Synthesis route 1

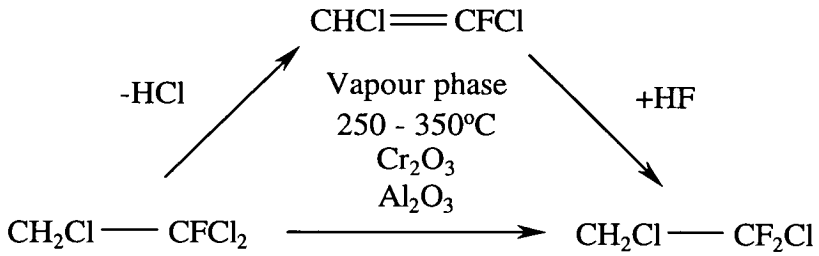
The starting material for this process is trichloroethene synthesised from methanol and sodium chloride. The first step in the reaction is the addition of hydrogen fluoride (HF) across the double bond (Equation 1.3). Of the two possible isomeric products, HCFC 131 (1,1,2-trichloro-1-fluoroethane) and HCFC 131a (1,1,2-tricholo-2-fluoroethane), it is found that HCFC 131 predominates. This is because the reaction follows Markovnikov's rule[27] which states: in the ionic addition of an acid to a carbon-carbon double bond of an alkene, the hydrogen of the acid attaches itself to the carbon atom that already holds the greatest number of hydrogens. The reaction takes place in the vapour phase at high temperature using an alumina or chromia catalyst. The specific conditions and catalysts used by each company are different and closely guarded commercial secrets.

Equation 1.3

$$CHCl{=}{=}CCl_2 + HF \xrightarrow[Cr_2O_3 \; Al_2O_3]{\text{Vapour phase } 250 - 350^{o}C} CH_2Cl{-}{-}CFCl_2 + HCl$$

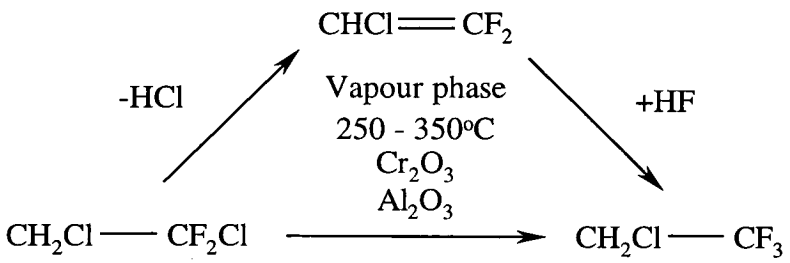
The second step in the process is the further addition of HF across the carbon-carbon single bond (Equation 1.4). The first stage in this mechanism is thought to be an elimination reaction where hydrogen chloride (HCl) is lost from HCFC 131 to form HCFC 1121. HF is then substituted across the carbon-carbon double bond, again following Markovnikov's rule, to form HCFC 132b.

Equation 1.4



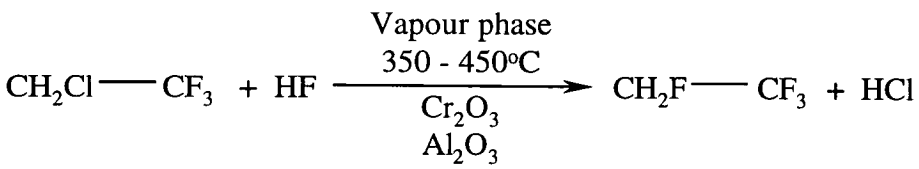
Step three involves a substitution of HF across the carbon-carbon single bond (Equation 1.5). The mechanism is again thought to be through the elimination of HCl to form HCFC 1122 followed by the addition of HF across the carbon-carbon double bond. The final product formed is HCFC 133a.

Equation 1.5



The final step in the reaction takes place in a separate reaction vessel from the first three steps (Equation 1.6). The same types of catalysts are used but the temperature is increased to 350 - 450°C. The reaction involves the substitution of chlorine by fluorine. This reaction takes place by nucleophillic substitution as chlorine is a better leaving group than fluorine.

Equation 1.6

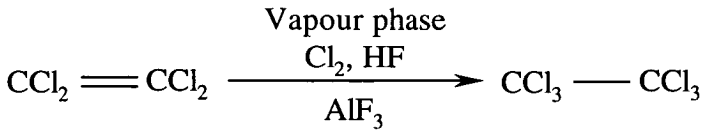


1.3.4 Synthesis route 2

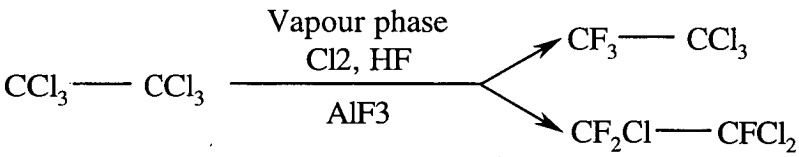
The starting material for this route of synthesis is perchloroethene. The first step in the synthetic pathway is to fully saturate the carbons with chlorine (Equation 1.7).

The reaction occurs in the vapour phase using chlorine gas with an AlF_3 catalyst. The mechanism of the reaction is thought to be a two step process: first, addition of a positive chloride ion to form an organic cation; then combination of this cation with a negative chloride ion. As both carbon atoms have chlorine substituted onto them the stereospecificity of the reaction is not a concern.

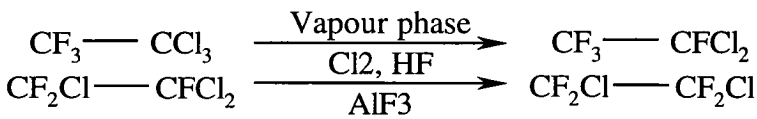
Equation 1.7



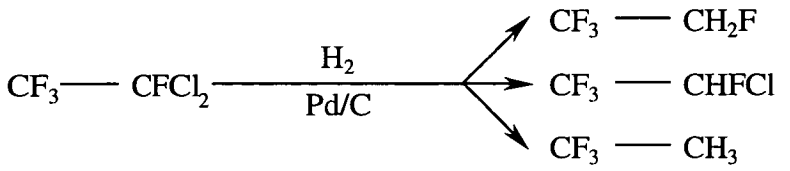
The second set of reactions are carried out in the same vessel and comprise of the sequential addition of fluorine atoms onto the carbons (Equation 1.8). The reaction is a nucleophillic substitution mechanism where chlorine is a better leaving group than fluorine. The reactions that occur in these steps are not stereospecific and lead to the formation of more than one isomer. The addition of one fluorine atom forms the only isomer CFC 111; addition of two fluorine atoms forms isomers CFC 112 and CFC 112a; addition of three fluorine atoms forms CFC 113 and CFC 113a; and final addition of four fluorines forms either CFC 114 or CFC 114a. CFC 114a is extracted and transferred to another vessel for the final stage of synthesis.



Equation 1.8



The final stage of the reaction is the removal of the two chlorine atoms to give HFC 134a. This is achieved by a two step reduction reaction of CFC 114a using hydrogen gas in the presence of a palladium/carbon catalyst (Equation 1.9). This reaction leads to the formation not only of the required HFC 134a, but also of high levels of the by-products HCFC 124 and HFC 143a (5-10% and 3-5 %, respectively).



Equation 1.9

1.3.5 Comparison of the two processes

Route 1 has a greater level of selectivity than that achieved with Route 2. HFC 134a can be over 99% pure at the final stage. This is a major advantage when compared with the 85-90% purity gained with Route 2. However, the two major impurities formed in Route 2, HCFC 124 and HFC 143a, have commercial applications and can therefore be used after separation from HFC 134a.

The major advantage of Route 2 over Route 1 was that the initial capital expenditure on setting up HFC 134a plants was comparatively small. The plants that had been previously used for the production of CFC compounds could be modified to produce HFC 134a using Route 2. Route 1 required new dedicated plants to be built in order to produce the quantities of HFC 134a required. This places a large initial financial strain on any company trying to produce HFC 134a.

1.3.6 Sources of contaminating by-products

To consider all the possible sources and possible types of contaminating by-products is beyond the scope of this work. Therefore, an overview of the major sources and types of by-products which may be present in HFC 134a will be detailed. Appendix B lists the by-products previously seen in samples of HFC 134a.

The first major area for possible contaminants is the raw materials used. In Route 1 trichloroethene is the starting material. Trichloroethene is synthesised by reacting methanol and sodium chloride. The sodium chloride is obtained from salt mines and will therefore contain small levels of sodium bromide. Unless lengthy and costly purification steps are taken before synthesis of trichloroethene, trace levels of bromine containing compounds will be present. Methanol will also contain

contaminants which will form by-products when used to synthesize trichloroethene. The reaction process to form trichloroethene will not be 100% efficient and so the starting material will contain levels of partly and fully chlorinated C_1 (*e.g.* CH_2Cl_2), C_2 (*e.g.* $C_2H_2Cl_2$), C_3 or longer chain carbons.

The isomeric selectivity of each reaction is important in forming the correct product at each step of the process. Route 1 shows high levels of selectivity in the first part of the process as it follows Markovnikov's rule of substitution, although there will be trace levels of unwanted isomers present. Route 2 shows more clearly the problems encountered when the reaction has a reduced level of specificity. Even though the conditions the reactions take place under and the catalysts used try to optimise the formation of the required products, large amounts of unwanted isomers are formed.

Contaminating by-products are also formed if a series of reactions does not go to completion or continues past the preferred final product. In the case of Route 1 the final product of HFC 134a may contain residuals of unreacted HCFC 132b. In route 2 the continual substitution of fluorine into HCFC 114a will form CFC 115 which may then be subsequently reduced to form HFC 125.

1.3.7 Sample purification

There are two main ways by which HFC 134a or its reaction intermediates can be purified[28]: i) by fractional distillation and/or, ii) by adsorption. Fractional distillation is a technique where the different boiling points of substances in a mixture are employed to allow separation. Crudely speaking, the components are separated by sequential boiling off the components and collecting the required purified distillates.

Adsorption of the different components onto solid surfaces can be used to remove impurities. The different affinities that the components have for solid phases, such as activated carbon, can be used to either extract by-products or to retain synthesis intermediates, leading to purification of the sample. Components may also be adsorbed using zeolites which work on the basis of polarity or on pore size.

Whichever method is used the degree of sample clean up will ultimately depend on the end users' requirements. There is little point in performing expensive clean up operations to produce a high purity level of HFC 134a >99.99% if the users' requirements only specify purity of >99% (*e.g.* high purity required for pharmaceuticals, lower purity required for air conditioners).

1.4 Sample analysis

In the work reported here, the characterisation of samples of HFC 134a was performed by measuring and identifying the levels of synthesis by-products present. This was done by separating the by-products from the major HFC 134a component using gas chromatography. The by-products were then detected using either flame ionisation detection of electron capture detection. The by-products were identified using gas chromatography-mass spectrometry. The chromatographic profiles produced by flame ionisation detection and electron capture detection were interpreted using chemometric analysis techniques.

1.5 Chromatography

Chromatography is the process by which a mixture of substances is separated into its component parts[29]. There are a number of different types of chromatography but they all work on the same principle of distribution of components between a mobile and stationary phase. The mixture is introduced into a constant flow of mobile phase which travels along in contact with the stationary phase. The different components in the mixture have different affinities for the stationary phase. Those with high affinities interact with the stationary phase and are retarded; those with lower affinities interact less with the stationary phase and travel more quickly. This leads to separation of the different components on the basis of their affinities for the stationary phase.

1.5.1 Gas Chromatography

Gas chromatography uses a gas as the mobile phase and either a liquid or solid as the stationary phase. The stationary phase can either be coated onto particles which are

placed into tubes (packed columns) or coated onto the walls of tubes (capillary columns)[30]. The mixture for separation is introduced into a continuous stream of gas (carrier gas) and carried along the tube, being separated on the basis of the components' affinity for the stationary phase.

The first publication describing gas chromatography was by Martin and James in *Biochem. J.*, **50**, 679 (1952). Thus gas chromatography as we know it today was developed in the late 1950s and early 1960s, mainly as packed column techniques. Capillary columns were developing at the same time[31] but it was not until the development of fused silica columns that problems of fragility and variability in column performance were overcome. The introduction of chemically bonded fused silica columns in the 1980s rejuvenated GC and considerably extended its range of applications.

The main limiting factor in GC is the requirement of the mixture to be volatalised into the gas phase. The temperatures required to do this are often in excess of the maximum operating temperatures of the columns used to separate the mixtures *i.e.* the sample is in gaseous state but the stationary phase has been stripped from the column. Even with columns of operating temperatures in the region of 400°C there is still the problem of the thermal shock experienced by the sample as it is volatalised. Compound derivatisation in some cases can provide an excellent method of making a compound more volatile to allow GC analysis. A common derivatisation method involves the acetylation of alcohols, phenols and amines using acetic anhydride in the presence of an acid catalyst (Equation 1.10). The obvious problem in derivatisation is that it introduces an extra step in the analysis which requires validation, especially when performing quantification work. The potential components in the different sample mixtures in the work performed here all have relatively low boiling points (48°C for HCFC 123)[32] and therefore are easily volatalised.

Equation 1.10

$$ArOH + (CH_3CO)_2O \longrightarrow CH_3CO_2Ar + CH_3COOH$$

1.5.1.1 Column performance

As with any analytical technique there are a number of parameters which need to be checked and optimised to give the best separation conditions possible when using GC. The following sections give details of the mathematical principles behind calculation of the column performance.

1.5.1.2 Separation efficiencies

The efficiency of a GC column can be expressed in terms of theoretical plates. This value can be considered as the number of times that a solute is desorbed and re-adsorbed between the stationary and mobile phases. This is a generalised term for what is in fact a dynamic process. The number of theoretical plates is governed by the retention time and is inversely proportional to the degree of peak broadening. The calculation of theoretical plates is performed using the measurements from Figure 1.2 substituted into either equation in Equation 1.11.

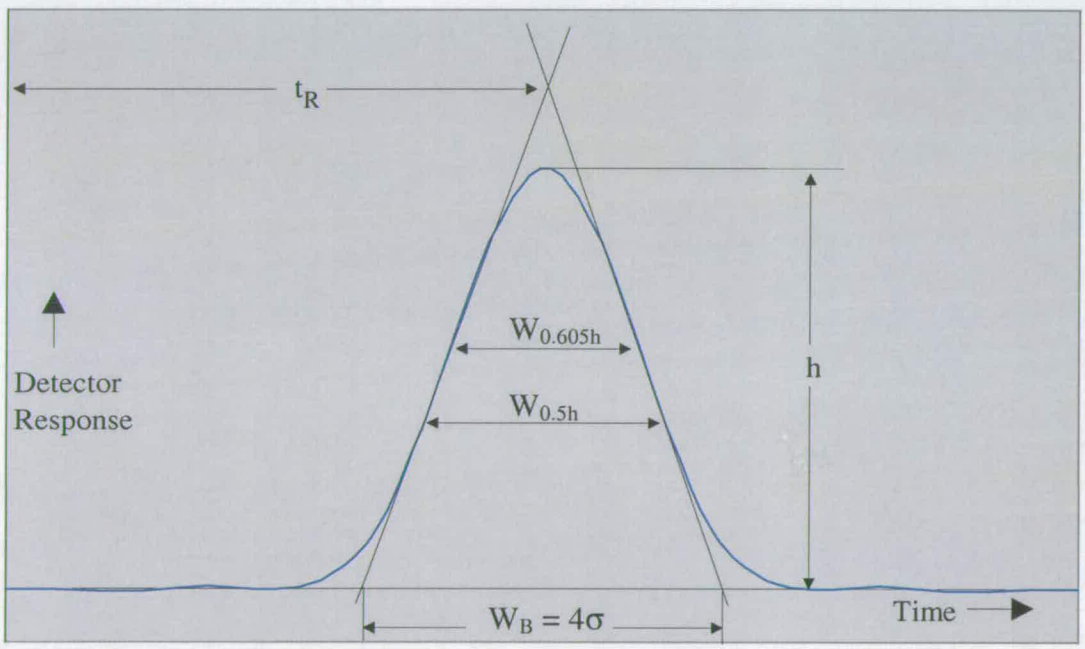


Figure 1.2: *Peak measurements for calculation of plate number*

$$N = 4\left(\frac{t_R}{W_{0.6065h}}\right)^2$$ **Equation 1.11**

$$N = 5.54\left(\frac{t_R}{W_{0.5h}}\right)^2$$

A more correct measure of the retention time and hence plate number is obtained by calculating the adjusted retention time, $t_{R'}$ (Equation 1.12). Non-retained components require a certain volume of gas to travel the length of the column, termed the dead volume. The time taken to travel the column is called the dead time t_M. The dead time is calculated as the retention time of an unretained component peak *e.g.* N_2O or methane (Figure 1.3).

$$t_{R'} = t_R - t_M$$ **Equation 1.12**

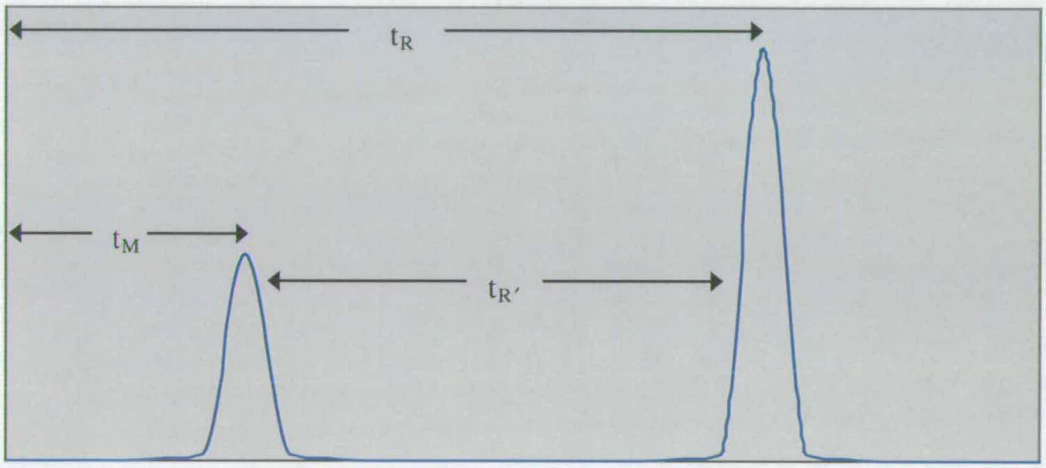


Figure 1.3: *Calculation of adjusted retention time*

The calculation using $t_{R'}$ gives a more accurate value for the number of theoretical plates, deemed the effective theoretical plate number (N). The higher the number of plates, the greater the ability of the column to separate two analytes in a mixture. Increases in column length increase the number of theoretical plates. However, at longer column lengths the pressure drops from the optimum carrier gas pressure which affects the peak shape *i.e.* a linear relationship of plate number to column length does not hold over longer columns. The efficiency of a column is normally quoted as the length of column occupied by one theoretical plate (H) which allows

easier comparison of efficiencies between columns of different lengths (Equation 1.13).

$$H = \frac{L}{N}$$

H = height per effective theoretical plate (HETP)
N = theoretical plate
L = Column length (m) **Equation 1.13**

1.5.1.3 Peak shape

The peak asymmetry factor (A_s) is a measure of the peak distortion, sometimes referred to as peak skew factor. Measurements are taken as shown in Figure 1.4 and the factor calculated using Equation 1.14[33].

$$A_s = \frac{CB}{AC}$$ **Equation 1.14**

Ideally values of very nearly 1.0 should be obtained. Values exceeding 1.2 indicate poor peak shape due to excessive amounts of peak tailing. Peak shape factor values of less than 1 indicate overloading of the column.

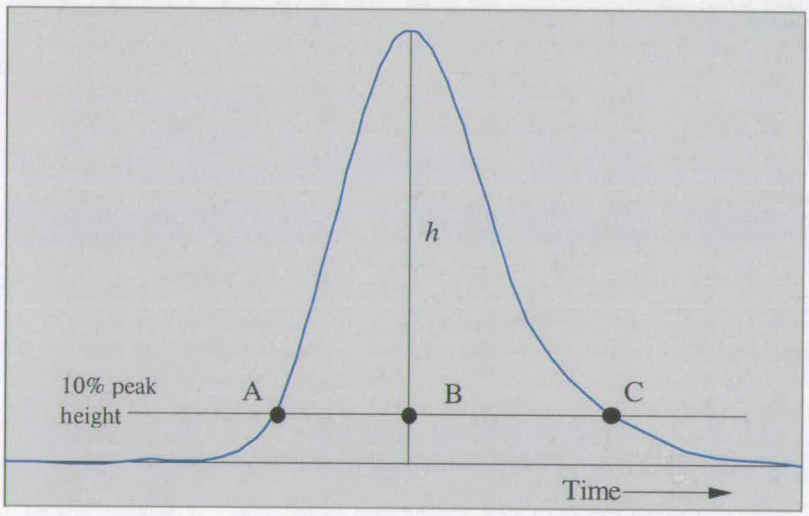


Figure 1.4: *Measurements for peak asymmetry calculation*

1.5.1.4 Carrier Gas

Choice of carrier gas type and operating pressure is vital to any gas chromatographic separation. Any gas to be used in capillary GC must be of the highest purity. This is achieved by using the manufacturers' highest purity gases and passing the gases

through a series of gas filters. These filters remove moisture, oxygen and, where possible, any interfering impurities (*e.g.* low molecular weight hydrocarbons).

The different performance characteristics of carrier gases can be evaluated by looking at the van Deemter plots[34]. Van Deemter graphs are generated by plotting the HETP against the average linear velocity for the same solute separated using the same column. Figure 1.5 compares the three carrier gases of hydrogen, helium and nitrogen. From the graphs it can be seen that as the molecular weight of the gas decreases, the optimum gas velocity increases. Also, the trough in the Van Deemter curves for lower weight molecular gases is considerably flattened. The practical consequence is that for lower molecular weight gases maximum efficiency occurs over a wider range of gas velocities and at higher gas velocities *i.e.* shorter chromatographic run times.

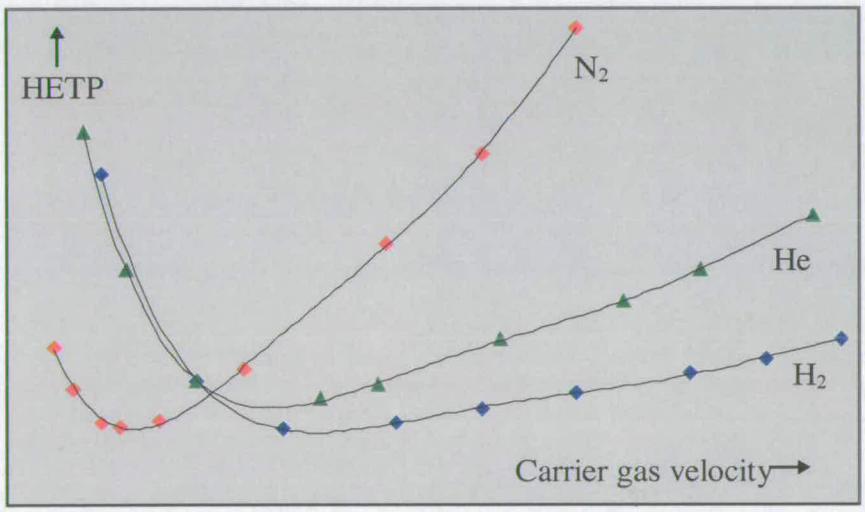


Figure 1.5 : *Comparison of HETP plots using different carrier gases*

Hydrogen is theoretically the best carrier gas. However, practical problems with leaks and the associated risk, while small, of explosion means that helium is usually the gas of choice. Consideration of pumping efficiency when coupling the GC to a mass spectrometry are also taken into account so that low molecular weight gases are preferred, as they are more quickly pumped.

1.6 Gas chromatography detectors

There are a number of different types of detector that are routinely installed on GC equipment[35]. The detectors all have their own merits and are employed for analyses dependent on their specific selectivity for the compounds under study. They all perform the same task of monitoring the GC column eluant and eliciting some form of electrical response as solutes elute. The detectors evaluated and employed in this work are the flame ionisation detector (FID) and the electron capture detector (ECD).

1.6.1 Flame ionisation detector

The FID is the most widely used of GC detectors[36]. It is the standard detector fitted to GC and shows sensitivity towards most compounds. It displays a wide linear range (10^7) for detector response compared with analyte concentration and has good reproducibility. The FID has been used in many studies involving halocarbons although the sensitivity towards fully halogenated hydrocarbons is somewhat lower than for non halogenated hydrocarbons[37].

The construction of a typical flame ionisation detector is shown in Figure 1.6. It consists essentially of a base in which the column eluant is mixed with hydrogen (40 ml min^{-1}), a polarised jet and a cylindrical electrode arranged concentrically with the flame. Air is supplied (400 mls min^{-1}) to the detector to support combustion. The assembly is contained in a stainless steel or aluminium body to which are fitted a flame ignition coil and electrical connections to the collecting electrode, and a polarising voltage to the detector jet. This polarising voltage sets up a potential difference between the jet and the collector. Ions formed due to the combustion of eluting analytes move to the collector electrode. The resulting ionisation current is then amplified and fed to a data system. The FID is heated to 275°C to prevent condensation of water generated by the flame and also to prevent any hold-up of solutes as they pass from the column to the flame.

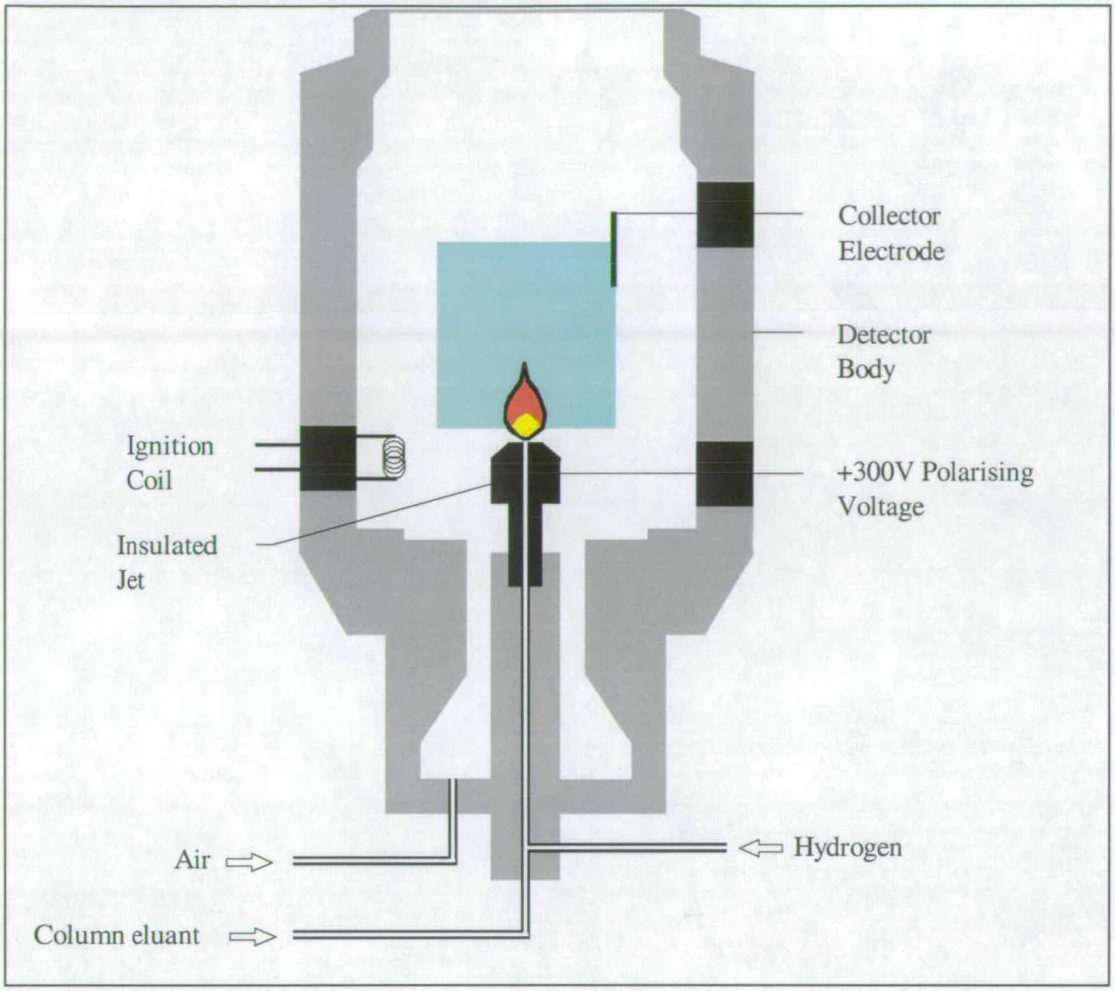


Figure 1.6: *Diagram of a typical FID*

1.6.2 Electron capture detector

The ECD was the first selective detector to be invented for GC[38] and was used as an extremely sensitive detector to measure atmospheric levels of chlorofluorocarbons. The internal chamber of the detector is lined with a radioactive ß-particular emitter (^{63}Nickel) contained in a sealed foil. The nitrogen make-up gas is ionised by the ß-radiation to generate free electrons which move quickly to the anode under the influence of the potential gradient before they can recombine with the nitrogen cations to form neutrals. This generates an ion current (Equation 1.15)

$$N_2 \longleftrightarrow N_2^+ + e^- \qquad \textbf{Equation 1.15}$$

Components capable of capturing electrons will form ions. These component ions can be neutralised by the nitrogen cation. The net effect is a decrease in total ion current. (Equation 1.16).

$$\text{Component} + e^- \longleftrightarrow \text{Component}^- \qquad \textbf{Equation 1.16}$$
$$\text{Component}^- + N_2^+ \longleftrightarrow \text{Component} + N_2^-$$

Therefore, the ECD is most sensitive for compounds which contain electronegative elements which allow the easy attachment of electrons[39], *e.g.* halogens. The ECD is extremely sensitive to chlorine containing molecules (detecting part per trillion concentrations in the atmosphere without the need for sample pre-concentration) but does have drawbacks in that it has a limited linear dynamic range (four orders of magnitude) and can give rise to an unstable baseline[40].

1.7 Mass Spectrometry

In terms of the amount of sample the mass spectrometer is generally thought of as the analytical instrument which can supply the greatest amount of information about a molecule[41]. Often both conformational and structural information can be obtained with as little as picogram amounts of material[41]. The continuing advances in instrument design and the coupling of mass spectrometers to separation techniques has led to the widespread application of the instrument in fields as varied as pharmaceuticals, petrochemicals and biotechnology. The mass spectrometer is and will continue to be an essential tool in modern analytical science.

1.7.1 History of mass spectrometry

The theory of mass spectrometry is based on the ideas of combination chemistry laid down by Dalton in the early nineteenth century. The idea that elements in a compound are bound in fixed proportions by weight allows for their measurement and conformation to be elucidated.

The developmental work to modern mass spectrometers can be traced back to J J Thomson at the turn of the century[42]. Thomson discovered electrons in 1897 whilst

conducting experiments to link matter with electricity. The experiments performed by Thomson showed the existence of a new particle with a specific mass and negative charge. This particle behaved like no other known atomic particle. Since electrons emerged from matter they must be part of the atom but the precise relationship remained unknown until just before the First World War.

Around the same time as Thomson was performing his work Goldstein was experimenting with low pressure discharge tubes fitted with a perforated cathode. He found that when a high potential was applied to the electrodes glowing beams of particles were observed passing through the holes into the chamber beyond. It was not until Wien in 1898 deflected these particles using magnetic and electric fields that they were shown to be beams of positively charged particles.

Further work by Thomson, looking at the particles left behind after the formation of electrons using positive ray parabolas, confirmed Goldstein's work that they were in fact positive charged species formed by the loss of one or more electrons. Thomson's continuing work led to the discovery of two stable isotopes of neon. The first true mass spectragraph was produced by Aston in 1919 whose instrument had the added ability over Thomson's of being able to focus as well as separate the beam of positively charged particles resulting in greater resolution.

Continuing work in the field showed that some atoms are capable of accepting electrons and so become negatively charged. It is the principle of 'charging' or ionising atoms that is the basis for mass spectrometry as these charged species can be acted upon by electrical or magnetic fields unlike their neutral counterparts.

More elaborate instruments were designed until, in the 1940s, the first commercially available mass spectrometer was designed for use in the petrochemical industry. Since then the use and application of mass spectrometers have increased rapidly because of improvements in mass analyser technology, advances in vacuum systems and rapid development of electronics and computers.

The modern mass spectrometer can be split into four main regions as displayed in Figure 1.7. The mass spectrometer requires a sample inlet system which may either be 'one off', *e.g.* direct probe, or continuous, *e.g.* gas chromatography inlet. The sample is then ionised by one of a number of techniques, *e.g.* electron impact, fast atom bombardment. The sample is then extracted into the analyser assembly where the ions are separated on the basis of their mass to charge (*m/z*) ratio. Finally the ions are detected and an amplified signal is sent to a suitable data handling system, usually a computer.

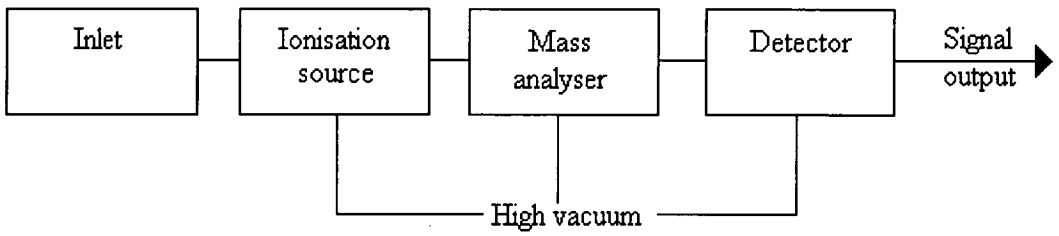


Figure 1.7: *Schematic of modern mass spectrometer design*

To function satisfactorily the mass spectrometer must operate under high vacuum (typically 10^{-6}torr) which is a problem when coupling the instrument to a continuous flow of sample *e.g.* liquid chromatography. A combination of rotary pumps, to rough pump, and oil diffusion and/or turbomolecular pumps, to fine pump, is employed to achieve the high vacuum required. It is the high rates of pumping that allows the GC eluant from a capillary column (in the order of 2mls min^{-1}) to be fed directly into the ionisation region without the requirement of any novel interface.

1.7.2 Sample Ionisation

There are a number of methods to achieve sample ionisation but by far the most common, and the technique used in this work, is electron impact ionisation (EI)[43]. The EI source (Figure 1.8) was first introduced to mass spectrometry by Dempster[44] and developed by Nier[45]. The principles of EI are well documented[46]. The ionisation region is constructed around the ion chamber into which vapourised sample is introduced. In the case of capillary GC the column terminates a couple of mm from the electron beam. The electron beam is formed by passing a current through the

filament, a tungsten or rhenium wire, which is maintained at 0-100V with respect to the ionisation chamber. The actual value of this difference is referred to as the electron energy (eV). By definition, 1eV is the energy imparted to one electron as it is accelerated through a 1V field. In practice the ionisation potential is held at 70V (70eV) as this is sufficient electron energy to give reproducible ionisation of molecules. The accelerated electrons travel across the ionisation chamber where they collide with an electrode called the trap. This is used to monitor the current travelling across the chamber.

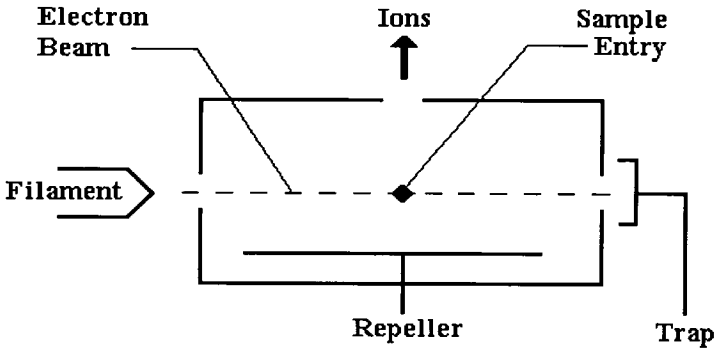


Figure 1.8: *Schematic of an electron impact ionisation source*

Some of the electrons travelling across the chamber collide with some of the column eluant resulting in either the sample molecules giving up an electron or capturing an electron (Equation 1.17). The formation of negatively charged ions is approximately $x10^2$ less likely to happen.

$$M + e^- \longrightarrow M^+ + 2e^-$$
$$M + e^- \longrightarrow M^-$$

Equation 1.17

The repeller plate is held at a small positive potential with respect to the ionisation chamber. The potential difference encourages any positively charged ions formed by electron/analyte collisions to migrate towards the mass analyser where they are accelerated and focused into the analyser assembly. Only relatively few of the analyte molecules entering the ionisation source collide and form ions, the amount of ionisation that occurs is described as the ionisation efficiency[47] and is usually in the order of 10^{-4}.

1.7.3 Mass Analysers

There are a number of different mass analysers used to measure the *m/z* ratio of ions all of which have been well documented in the literature. The mass analysers in the mass spectrometers used in this work are of two kinds. The Kratos instrument uses a reverse geometry magnetic sector and electric sector[48] and the Finnigan mass spectrometer employs a quadrupole mass filter[49].

1.7.3.1 Magnetic Sector

The magnetic sector analyser is in the shape of a sector of a circle. Ions entering this field will travel in a circular direction with the radius of the circle defined by the momentum of the ion. Therefore ions of the same mass, kinetic energy and charge are focused to one point. In practice ions of the same charge do not always have the same kinetic energy, usually due to the ionisation process, and therefore have what is termed energy spread. The consequence is that an instrument with only a magnetic sector, described as single focusing, has limited resolving power.

1.7.3.2 Electric sector

An electric sector instrument consists of two metal plates that form a sector of two concentric cylinders. Each plate has an equal but opposite potential applied to it. An ion entering the sector is subjected to a force generated by the electric field and the ion describes an arc of a circle. The radius of flight is dependent only on the kinetic energy of the ion. Thus ions are separated on the basis of their kinetic energies.

1.7.3.3 Double focusing instrument

By placing a magnetic sector and electric sector in series the electric sector can overcome the problems of energy spread encounted with the magnetic sector. The outcome is an instrument with greatly enhanced resolving power (Figure 1.9). The mass analysers can be assembled electric-magnet sectors (forward geometry) or magnet-electric sectors (reverse geometry).

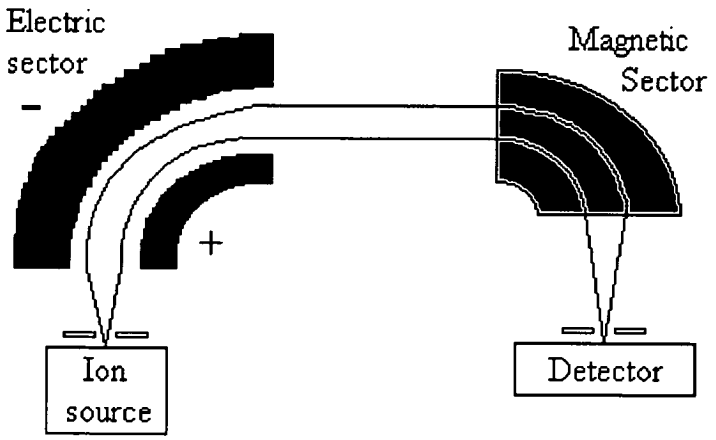


Figure 1.9: *Double focusing forward ion optics*

1.7.3.4 Quadrupole analyser

The quadrupole mass analyser consists of four precisely machined parabolic rods symmetrically arranged (Figure 1.10). Opposite rods are electrically connected. To one pair of rods a positive DC current is applied and to the other pair a negative current is applied. Superimposed over one pair of rods is an AC current, in the radio frequency, called an Rf voltage.

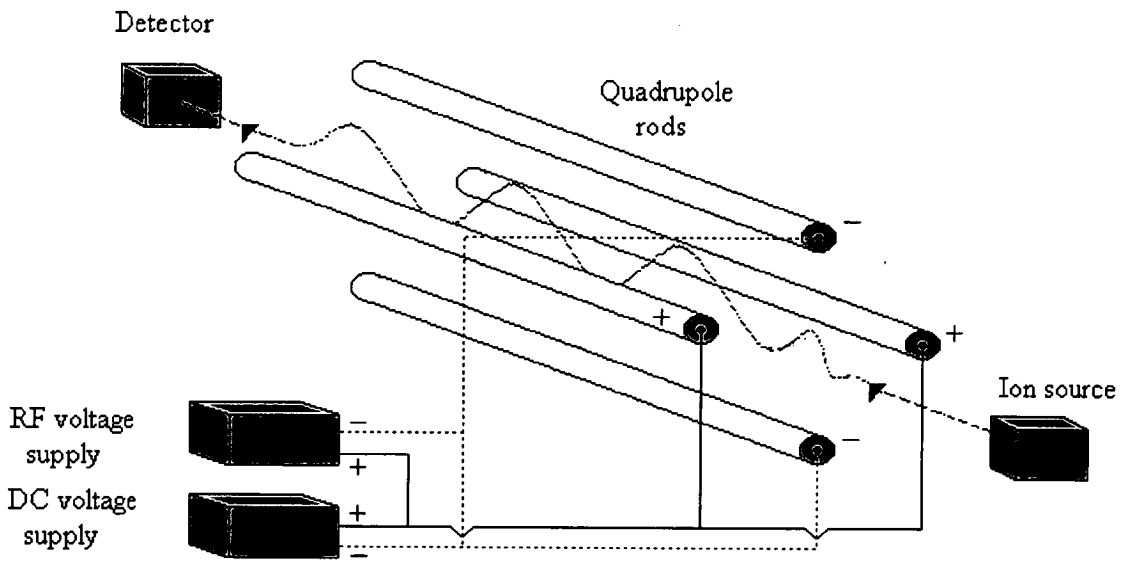


Figure 1.10: *Quadrupole ion optics*

The DC and Rf voltages are kept at a constant ratio and any ions entering the electric field will undergo oscillations. At a given value for the voltages, ions of a specific

m/z ratio will undergo bounded oscillations and be transmitted through the analyser assembly. Ions of different *m/z* ratio will undergo unbounded oscillations and will not be transmitted. By scanning across a set ratio of DC/rf voltages, ions can be transmitted sequentially on the basis of their *m/z* ratios.

1.7.4 Ion detection system

The ions exiting the mass analyser can be detected in a number of different ways[50]. By far the most common, and the system used in the mass spectrometers in this work, is the continuous dynode electron multiplier. Ions exiting the mass analyser strike the curved surface of the dynode and lead to the formation of secondary particles, electrons, which in turn continue down the dynode striking the walls forming more secondary particles in a cascade effect. The net result is an amplification (typically 10^6) of current to a level which can be detected. The signal is then transferred to a data handling system (usually a computer) via the appropriate analogue to digital converter.

1.8 Signal processing

The digital signals from the GC detectors and mass spectrometer are sent to data handling systems. These signals are displayed as chromatographs, in the case of the GC detectors, and as ion chromatographs and mass spectra, in the case of the mass spectrometer. The signal can either be processed automatically, under the constraints of pre-defined parameters, or manually. Automated processing is less time consuming but care should be taken as often information may be ignored or overlooked due to the inflexibility of the process. Whichever method is used there are a number of considerations to be taken into account when analysing the raw data. The following sections highlight the areas to be considered when extracting the information required from the raw data.

1.8.1 Background noise levels

Signal output from detectors is subject to levels of background noise[51]. The sources of this noise are varied *e.g.* electrical noise, detector stability. It is not normally important to identify these sources if the instrument is working correctly. It is however, vital to quantify the levels of noise to monitor the correct operation of the instrument and for use in peak identification. Noise levels are monitored by measuring the total height of signal (see Figure 1.11) at a steady state. Noise levels should be calculated across the chromatogram as they may alter with experimental conditions *e.g.* when using programmed column temperature gradients.

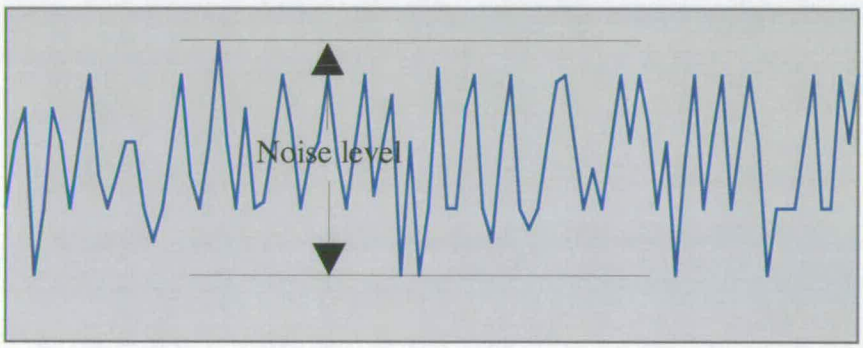


Figure 1.11: *Measurement of chromatographic background noise level*

1.8.2 Baseline drift

The signal output from a detector should ideally be a constant in time (± noise). In practice, the absolute value for the signal may increase or decrease with time (Figure 1.12)[58]. Drift is often present during routine operation, although this may indicate a fault with the detector *e.g.* detector not fully warmed up.

The temperature programming of the GC affects the temperature of the eluting carrier gas which in turn will affect the absolute baseline value of the detector. Changes in the absolute value of the baseline will affect the values obtained for peak heights/area. To compensate for any drift observed the baseline is not considered as a static value and peak calculations are adjusted accordingly.

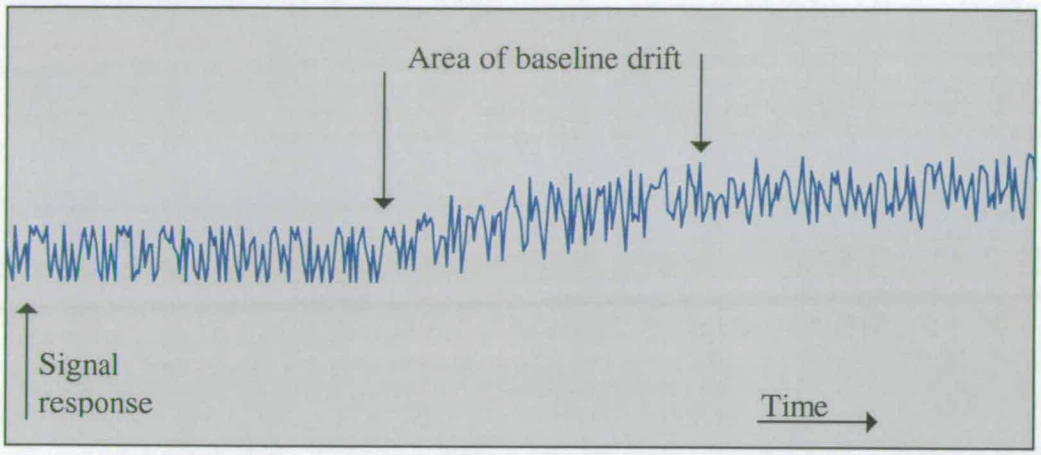


Figure 1.12: *Diagram to illustrate baseline drift*

1.8.3 Peak Definition

There has been a great deal of discussion on the precise mathematical definition of what constitutes a peak[52]. All discussions, however, agree that certain criteria must be met to distinguish a response on a chromatographic trace due to a component from a response because of noise or an electrical spike.

A response in a chromatogram will be considered a true peak in this work if the height of that peak exceeds a ratio of 3:1 signal to noise. That is, peaks must be at least three times the height of the surrounding noise level before they will be accepted as a response to a component. This value is described as the limit of detection (LOD). A second criteria to consider is the limit of quantification (LOQ). This value refers to the peak height required to give an accurate representation of the amount of a component present. This value is more susceptible to error as small variations in peak height value can greatly affect the result obtained for quantification. For this reason the LOQ is set a signal to noise ratio of 5:1.

Electrical spikes can be filtered out of the data by setting a minimum time value for peak width. Therefore no response above LOD will be considered as a true peak unless the width of that peak is greater than 5 seconds.

1.8.4 Data format

Transfer of data from analytical equipment to the statistical analysis packages can often be problematic. Different instrument companies use different computer systems and even if the computer systems are the same the file formats they use may not be. Considering every point in the chromatograph is not always necessary and the number of data points can be reduced to the values representing the peaks present (*i.e.* peak heights). Therefore the data from a chromatographic analysis can be displayed in a table, where the different samples analysed are listed down the table, the peaks seen listed across the table, and the values for those peaks entered into the body of the table (Table 1.1). The data is then in a format where it can be transferred into and between statistical packages.

Table 1.1 : *Example of tabulated chromatographic data*

	Peak 1	Peak 2	Peak 3
Sample 1	$Value_{peak1\ sample1}$	$Value_{peak2\ sample1}$	$Value_{peak3\ sample1}$
Sample 2	$Value_{peak1\ sample2}$	$Value_{peak2\ sample2}$	$Value_{peak3\ sample2}$
Sample 3	$Value_{peak1\ sample3}$	$Value_{peak2\ sample3}$	$Value_{peak3\ sample3}$

References

[1] Wayne, R.P., *Chemistry of Atmospheres*, Clarendon Press Oxford, 160 (1991)

[2] Lovelock, J.E., *Nature*, **230**, 379 (1971)

[3] Spauschus, H.O., *Rev. Int. Froid*, **11**, 389-392 (1988)

[4] Lovelock, J.E. and Maggs, R.J., *Nature*, **241**, 194-196, (1973)

[5] Molina, M. J. and Rowland, F. S., *Nature*, **249**, 810-812, (1974)

[6] Thrush, B.A., *Rev. Porg. Phys.*, **51**, 1341 (1998)

[7] Wayne, R.P., *Chemistry of Atmospheres*, Clarendon Press Oxford, 121-123 (1991)

[8] Webster, G., *The Ozone Layer*, Department of the Environment, 94 EP 253 (1994)

[9] Klingman, L.H. and Zheng, P.S., *J. of Investigative Dermatology*, **100**, V2, p194-199, (1993)

[10] Farman, J.C., Gardiner, B.G. and Shanklin, J.D., *Nature*, **315**, 207-210, (1985)

[11] *Montreal Protocol on Substances that Deplete the Ozone Layer-Final Act.*, United Nations Environmental Programme, UNEP, (1987)

[12] Spauschus, H.O., *Rev. Int. Froid*, **11**, 389-392 (1988)

[13] Henrici R. and Preisegger E., *Int. J Refrig.*, **15,** 6, 326-331 (1992)

[14] Fisher, D.A., Hales, C.H., Filkin, D.L., Ko, M.K.W., Sze, N.D., Connell, P.S., Wuebbles, D.J., Isaksen, I.S.A. and Stordal, F. *Nature*, **344**, 508-512 (1990)

[15] Fisher, D.A., Hales, C.H.,Wang, W, Ko, M.K.W., Sze, N.D, *Nature*, **344**, 513-516 (1990)

[16] Kanakidou M., Dentener F.J. and Crutzen P.J., *J of Goephysical Research*, **100**, D9, 18,781-18,801 (1995)

[17] Wallington, T.J., Hurley, M.D., Fracheboud, J.M., Orlando, J.J., Tyndall, G.S., Sehested, J., Mogelberg, T.E. and Nielsen, O.J., *J. Phys. Chem.*, **100**, p18116-18122, (1996)

[18] Bowden, D.J., Clegg, S.L. and Brimblecombe, P., *Atmos Envir.*, **11**, 819-828, (1977)

[19] Visscher P.T., Culbertson C.W. and Oremland R.S., *Nature,* **369**, 729-731, (1994)

[20] Thompson R., Stewart K.M. and Gillings E., *AEFAS Rep. No. SP91-18.8*, (1994)

[21] Franklin, J., *Chemosphere*, **27**, 1565-1601 (1993)

[22] Schwarzbach, S.E., *Nature*, **376**, 297-298, (1995)

[23] Hartmut F., Klein A. and Renschen D., *Nature,* **382** 34 (1996)

[24] Wujcik C.E., Zehavi D. and Seiber J.N., *Chemosphere,* **36** N6 1233-1245 (1998)

[25] Lee, S.A. and Robinson, G.E., *Process Development: Fine Chemicals from Grams to Kilograms*, Oxford University Press, Oxford, (1995)

[26] Banks, R.E., Smart B.E. and Tatlow J.C., *Organofluorine Chemistry Principals and Commercial Apllications*, 159-175 Plenum Press, New York and London

[27] March, J., *Advanced Organic Chemistry: Reactions, Mechanisms and Structure.* 4th ed., John Wiley and Sons. New York, p750, (1992)

[28] *Personal Communication*, A Handley, ICI C&P, Runcorn, England

[29] IUPAC Analytical Division Commission on Analytical Nomenclature, *Pure and Appl. Chem.*, **37**, 447, (1974)

[30] Braithwaite, A. and Smith, F.J., *Chromatographic Methods*, 156-162 Chapman and Hall, London, (1985)

[31] Golay, M.J.E., *Gas Chromatography*, (Hrsg), Butterworth, London, p 36-55 (1958)

[32] Zeeuw, J., Zwiep, D. and Marinissen, J.W., *International Laboratory*, **September**, 12J-12P, (1996)

[33] Dyson, N., *Chromatographic Integration methods*, RSC Chromatography Monographs, Royal Society of Chemistry, p6, (1992)

[34] van Deemter, I.J., Zuiderweg, F.J. and Klinkenberg, A, *Chem. Eng. Sci.*, **5**, 271, (1956)

[35] Bachmann, K. and Reinke, F.J., *Journal of Chromatography*, **323**, p323-329, (1985)

[36] Eiceman, G.A, Hill, H.H. and Davani, B., *Anal. Chem.*, **66,** 621R-633R, (1994)

[37] Schomburg, G., *Gas Chromatography*, VCH Publishers, New York, p116-117, (1990)

[38] Lovelock, J.E. and Lipsky, S.R., *J. Amer. Chem. Soc.*, **82**, p431, (1960)

[39] Deveaux, P., and Guichon, G., *J. Gas Chromatogr.*, **5**, p314, (1967)

[40] Fowlis, I.A., *Gas Chromatography*, 2nd ed., John Wiley and Sons, Chichester, England, p146-148, (1995)

[41] Fifield, F.W. and Kealey, D., *Principles and Practice of Analytical Chemistry*, 4th ed., Blackie Academic and Professional, London, p436, (1995)

[42] Griffiths, I.W., *Rap. Comm. Mass Spec.*, **11,** 2-16, (1997)

[43] *Eight Peak Index*, 4th ed., Royal Society of Chemistry, Cambridge, (1992)

[44] Dempster, A.J., *Phys. Rev.*, **18**, p415, (1921)

[45] Nier, A.O., *Rev. Sci. Instrum.*, **18**, p398, (1947)

[46] Chapman, J.R., *Practical Organic Mass Spectrometry*, 2nd ed., John Wiley and Sons, Chichester, England, p1-4, (1993)

[47] Ashcroft, A.E., *Ionization Methods in Organic Mass Spectrometry*, RSC Analytical Spectroscopy Monographs, Cambridge, (1997).

[48] Johnson, E.G. and Nier, A.O., *Phys, Rev.*, **91**, 10, (1953)

[49] Campana, J.E., *Int. J. Mass. Spec. Ion. Phys.*, **33**, 101, (1980)

[50] Kurz, E.A., *American Laboratory*, March, (1979)

[51] Schomburg, G., *Gas Chromatography*, VCH Publishers, New York, p166-167, (1990)

[52] MacDougall (Chairman) *et al.*, *Anal. Chem.*, **52**, 2242, (1980)

Chapter 2

Chemometric Analysis

2.0 Introduction to chapter

Chemometrics is a term used to describe the application of mathematics to the investigation of data sets formed by chemical analyses[1]. Since the term's first usage, the range of mathematical processes it is deemed to include has broadened. This has led to the term being used to encompass a large array of mathematical processes applied to aid the understanding of analytical chemical data. The areas in which chemometric techniques are applied are also increasing rapidly[2].

Many analytical techniques employed in modern chemistry involve the measuring of not one but multiple parameters. The results for an infra-red spectrum are not described by one absorption peak but by many[3]. In fact, as the band width of wavelength measured becomes narrower the number of data points becomes huge. The comparison of these spectra becomes time consuming, mundane and often a difficult task for a highly trained scientist. It is here that the field of multi-variate statistics can be applied. These processes allow the results of analyses to be compared using well defined mathematical rules. The mathematical rules can be used to investigate data quickly and, in many cases, in new ways. The consequence is a set of techniques which speed up and aid the more complete interpretation of analytical data.

The chemometric techniques employed in this work were in the field of pattern recognition[4]. The ultimate goal of this thesis was to characterise samples HFC 134a on the basis of levels of by-products of synthesis. Similar samples were grouped and separated from dissimilar samples. Each sample was described by a series of values for the levels of by-products of synthesis present. Each by-product was considered as a separate variable.

The data sets formed by the analyses of HFC-134a samples were described by a series of variables where each variable represents the presence or absence of any by-product. The value for each of these variables was the measure of the quantity present of each by-product in the form of peak height/area detector response. The

samples were characterised on the basis of the levels of the by-products present. Therefore if a sample of HFC-134a did not contain a certain by-product the value given will be zero, as the absence of any by-product was just as important as its presence.

2.1 Introduction to multivariate mathematics

The following section gives an overview of the mathematics and principles which will be employed in the chemometric analyses performed in this work. It is not meant as a comprehensive review of statistics and multivariate statistics used but more as an introduction.

2.1.1 Normal distribution

Any set of analytical measurements, no matter how carefully obtained, will be subject to random errors and natural variation[5]. The value obtained for a measurement will vary if repeated. This variation can often be approximated using the normal distribution (sometimes called the Gaussian distribution)[6]. The mathematical model describing the normal distribution of a set of measurements is shown in Equation 2.1 where μ is the mean value of the data and σ is the deviation of points around the mean. The characteristic bell shape of the standardised normal curve (in which $\mu = 0$ and $\sigma = 1$) is displayed in Figure 2.1.

$$f(x) = \frac{1}{\sigma\sqrt{(2\pi)}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right] \qquad \textbf{Equation 2.1}$$

Equation 2.1 is an idealised one based on an infinite number of measurements. In practice the actual number of sample measurements will be finite, n, and the descriptive statistics will be estimates based on these measurements. The estimates for mean, variance and standard deviation are given by $\bar{x}$, s^2 and s, respectively, and calculated from Equations 2.2-2.4.

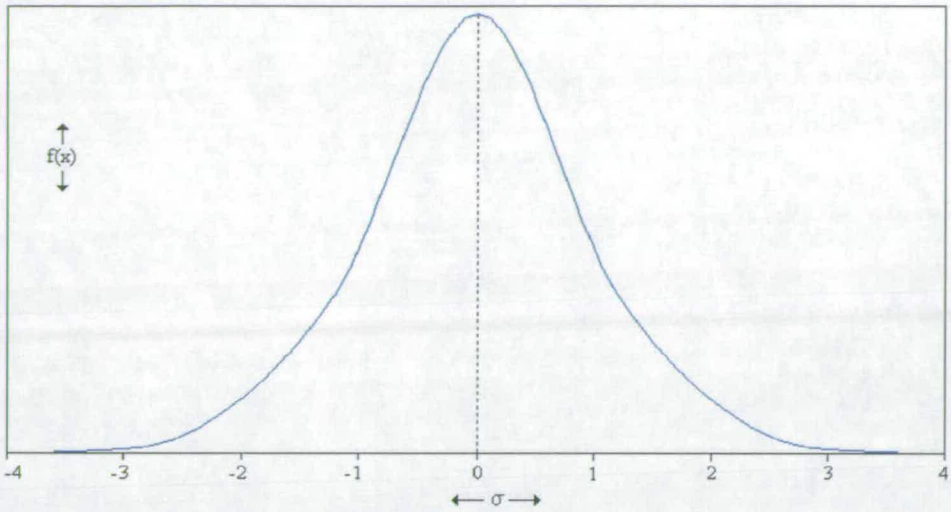


Figure 2.1: *Standardised normal distribution curve.*

$$\bar{x} = \sum_{i=1}^{n} x_i / n \qquad \textbf{Equation 2.2}$$

$$s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 / (n-1) \qquad \textbf{Equation 2.3}$$

$$s = \sqrt{(s^2)} \qquad \textbf{Equation 2.4}$$

These statistics can reduce single variate data into key values to allow data sets to be assessed easily. For most applications the mean and standard deviation are quoted to describe the data although a further value of the relative standard deviation can be calculated and quoted (Equation 2.5). This value presents the amount of spread of the data as a percentage of the mean and gives a less subjective value when compared with *s*.

$$\%RSD = 100s / \bar{x} \qquad \textbf{Equation 2.5}$$

2.1.2 Significance tests

Considering the normal distribution equation it can be calculated that ≈68.2% of results will lie within ±1μ of the mean value, ≈95.4% of results within ±2μ and ≈99.7% within ±3μ if the parent population of data follows the normal distribution.

A new value (x) can be compared with the expected parent population mean (μ_0) to see if the new value appears to be drawn from the expected parent population.

In statistical terminology a null hypothesis[7] is formed which states: the new value does not significantly differ from the parent population apart from random error. Limits are set as to what is significant *e.g.* within $\pm 2.58\sigma$ or 99%. The test statistic (z) is then calculated (Equation 2.6) for the new sample. The z statistic is then compared with a table of results listing values of z which fall within the different areas of confidence, the critical region. If the new calculated value of z lies within this region the null hypothesis can be accepted and the value belongs to the parent population. If the value for z lies outside this region the null hypothesis is rejected and the new value can be considered probably not to be part of the parent population.

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \qquad \textbf{Equation 2.6}$$

The above test is based on the idealised circumstances that all the data points for the parent population are known. In most analytical circumstances only a proportion of the parent population is known. In practical cases the significance test is performed using estimates of μ and σ and a probability spread wider than the normal distribution, *e.g.* Student's t-distribution[8]. The practical difference is the need to specify the number of degrees of freedom when comparing tabulated significance values with those calculated, where 'degrees of freedom' refers to the number of samples measured to build up the parent population.

The *F*-test is a further statistical test which compares the ratio of variances between two populations of observations to see if they belong to the same parent group[9]. If two random sets of populations are taken from a normal distribution of data and the s^2 values calculated, s_1^2 and s_2^2, then the ratio s_1^2/s_2^2 should be close to unity. By repeatedly taking different groups of the normal parent populations and plotting the ratio of variances the *F*-distribution curve will be formed. As with the Student t-test this distribution curve can then be used to place confidence levels on the spread of sets of data when comparing groups of observations.

2.1.3 Analysis of variance

Analysis of variance is a statistical test which compares a number of samples on the basis of a measured variable. That is, if six samples are measured five separate times can these values be investigated to see if the samples are different or if they belong to the same parent population? The main aim with this type of analysis is to separate the variation within a set of measurements of the same sample (*i.e.* random error) from the variation that may occur between samples. Experimental design plays a key role in reducing any effects from systematic errors, *e.g.* random analysis order. The statistical analysis of the group of samples is performed using an Analysis Of Variance (ANOVA) table[10](Table 2.1).

Table 2.1: *Commonly used layout for ANOVA table and calculation of F-test.*

Source of variation	Sum of squares	Degrees of freedom	Mean squares	*F*-test
Among samples	SS_A	m-1	s_A^2	s_A^2/s_W^2
Within samples	SS_W	N-m	s_W^2	
Total variation	SS_T	N-1	s_T^2	

The analysis of variance table allows the variance among samples to be compared with the variance within samples. If this value is significant, as calculated in an *F*-test, it can be said that the different groups of samples belong to different parent populations *i.e.* they are significantly different.

Total variance within all the samples and replicates is calculated using Equation 2.7.

$$s_T^2 = \sum_{j=1}^{m} \sum_{i=1}^{n} (x_{ij} - \bar{x})^2 (N-1) \qquad \textbf{Equation 2.7}$$

where $x_{i,j}$ is the i^{th} replicate of the j^{th} sample. The total number of analyses is N and n is the number of replicates per sample. Equation 2.7 can be rearranged to give the sum of the squares for the total variation, SS_T (Equation 2.8)

$$SS_T = \sum_{j=1}^{m} \sum_{i=1}^{n} x_{ij}^2 - \left[\sum_{j=1}^{m} \sum_{i=1}^{n} x_{ij} \right]^2 / N \qquad \textbf{Equation 2.8}$$

The variance among the different samples is obtained from SS_A (Equation 2.9)

$$SS_A = \sum_{j=1}^{m} \left[\sum_{i=1}^{n} x_{ij} \right]^2 / n - \left[\sum_{j=1}^{m} \sum_{i=1}^{n} x_{ij} \right]^2 / N \qquad \textbf{Equation 2.9}$$

The value for the within samples sum of squares is obtained from the difference between the total variance and variance among samples (Equation 2.10)

$$SS_W = SS_T - SS_A \qquad \textbf{Equation 2.10}$$

The *F*-test value can then be calculated for the among sample variance and between sample variance. This value is compared with table results to check for significance levels *i.e.* if the variance between samples is within limits of variance throughout the samples then they all originate from the same parent population.

2.1.4 Mulitivariate data

The techniques discussed thus far have been concerned with describing and comparing samples which are characterised by one variable. The following section discusses the statistical methods applied to multivariate data.

Table 2.2 displays the data obtained from the GC-FID analysis of five different standard C_1-C_4 hydrocarbon mixes (mV height responses). The data in Table 2.2 can be arranged in an array of size n x m where n is the number of samples or objects analysed and m is the number of variables measured, n = 5 and m = 4 for the displayed data set. The array is described as the data matrix and is dealt with as an entity rather than viewing each element of the matrix singularly. The data matrix is identified using a capital bold letter *e.g.* **X**. Each value within the data matrix is described as a variate and identified as $x_{i,j}$ where i denotes the row and j denotes the column. Single variate statistics can be applied to columns in order to gain information about any one variable but it is the interactions between the columns/variables which provides the most useful information. The measure of interactions between variables is described as their covariances[11].

Table 2.2: *Data set for the GC-FID analysis of five standard C_1-C_4 hydrocarbons.*

	Methane	Ethane	Propene	Butane
Sample 1	3.94	9.80	12.63	5.48
Sample 2	2.10	5.08	6.92	9.28
Sample 3	3.21	8.06	10.61	7.73
Sample 4	4.78	11.71	14.64	5.06
Sample 5	2.02	5.01	6.92	5.72
Mean	3.21	7.93	10.34	6.65
Variance	1.41	8.61	11.80	3.21

2.1.5 Covariance and correlation

Variance is described by the spread of data for a single variable about a mean. In the same way covariance is described by the mutual spread of data describing two variables about a common mean[12]. The equation used to calculate variance (Equation 2.3) can be extended to calculate covariance. Matrix notation is applied to provide a short hand representation of what would become unwieldy equations using conventional algebraic terms (Equation 2.11).

$$s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 / (n-1) \qquad \textbf{Equation 2.3}$$

If $x_d = (x_i - x)$ then the matrix $\mathbf{x_d}$ would be a column vector representing the values of $(x_i - x)$ for i = 1 to n. Equation 2.11 is the matrix notation of Equation 2.3.

$$s^2 = \sum x_d^2 / (n-1) \qquad \textbf{Equation 2.11}$$

where $\Sigma(\mathbf{x_d})^2$ is the same as $\mathbf{x_d}^T * \mathbf{x_d}$ where $\mathbf{x_d}^T$ is the transpose of $\mathbf{x_d}$ *i.e.* (row vector)* (column vector)

The numerator in the equation is the sum of the squares and in covariance the analogous quantity is the sum of the products (SP), where the SP is calculated between two variables (Equation 2.12)

$$SP_{jk} = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) \qquad \textbf{Equation 2.12}$$

where x_{ij} is the i^{th} element of variable j
x_j is the mean of variable j
x_{ik} is the i^{th} element of variable k
x_k is the mean of variable k

If $\mathbf{X_d}$ is the matrix which the describes the data set for the variables j and k the equation can be written in matrix form. As with variance the sum of the products is

divided by (n-1) to give the covariance between two variables. The results obtained for covariance are in the form of a covariance matrix which is square and symmetrical about its diagonal as the covariance of j to k is the same as k to j.

Table 2.3 shows the covariance calculations performed on methane and ethane from the hydrocarbon data in Table 2.2. The values obtained for covariance are difficult to interpret as the effect of the magnitudes of the original measurements effects the magnitude of the final result. A better representation of covariance is to use the correlation coefficient (r). This value takes into account the differences in magnitudes of variables. The correlation coefficient is calculated by dividing the covariance by the product of the variables standard deviations. The correlation coefficient lies between -1 and 1. Where values close to -1 show a high level of negative correlation, values close to 1 show high positive correlation and values near zero show the variables are linearly independent. Table 2.4 shows the results for the correlation coefficients for the hydrocarbon data.

Table 2.3: *Covariance calculation between methane and ethane values (Table 2.2)*

	Methane x_j	Ethane x_k	$(x_i - x_j)$	$(x_i - x_k)$	$(x_i - x_j)(x_i - x_k)$
	3.94	9.80	0.73	1.87	1.37
	2.10	5.08	-1.11	-2.85	3.16
	3.21	8.06	0.00	0.13	0.00
	4.78	11.71	1.57	3.78	5.93
	2.02	5.01	-1.19	-2.92	3.47
Average	3.21	7.93			
Sum	16.05	47.59			
Variance	1.412	8.612			

$SP_{methane, ethane} = 13.94$
covariance = 2.79

Table 2.4: *Correlation coefficients for hydrocarbon data*

	Methane	Ethane	Propene	Butane
Methane	1	0.800	0.798	-0.485
Ethane	0.800	1	0.800	-0.488
Propene	0.798	0.800	1	-0.489
Butane	-0.485	-0.488	-0.489	1

2.1.6 Multivariate normal

Most multivariate statistical methods are based on the assumption of a multivariate normal distribution for each population sampled. The exact definition of the equation for the multivariate normal distribution is complicated and difficult to visualise for more than two variables. However, the important factor in the equation is a quadratic form of the matrix product. This expands to give values which lie on the surface of an ellipse which is the characteristic shape of the multivariate normal, as the bell shape is characteristic of single variate statistics[13].

2.1.7 Combining variables

A useful technique in chemometrics is the consideration of linear combinations of the original variables. The objective is to describe the original data in terms of new variables where the new variables describe more of the original variance in fewer variables. That is, instead of ten variables describing the data it may be possible to describe the majority of the variance in the data using two or three variables. The number of total variables needed to describe the data remains the same but most of the new variables are describing residual amounts of variance *i.e.* not required to study the data. The combining of variables is performed using formal established mathematical techniques[14] which were first applied to chemical data by Malinowski around 1960.

2.1.7.1 Linear combinations of variables

Linear combination of variables involves the combination of variables to form a new variable. Consider the points plotted on the bivariate axes in Figure 2.2. These points represent measurements of two variables for each object. The distribution of these objects over each variable is represented as bar charts on each axis. By plotting a new axis, Var 3, and projecting the objects onto this axis it can be seen that more variance is described by the new axes *i.e.* the spread of objects along the axis is greater than either Var 1 or Var 2. This means that the new variable describes more of the original variance in the data than either Var 1 or Var 2.

The new variable can be defined as the sum of the original variables (Equation 2.13). The actual values for a and b are arbitrary but are usually defined as the normalised linear combination coefficients so that the new variables scale to the same range as the original axes. The power of this technique lies in finding the linear combinations of variables which describe the maximum amount of variance possible *i.e.* allow the reduction of the dimensionality of the data[15]. This technique is described in the following section.

$$Var3 = a.Var1 + b.Var2 \qquad \textbf{Equation 2.13}$$

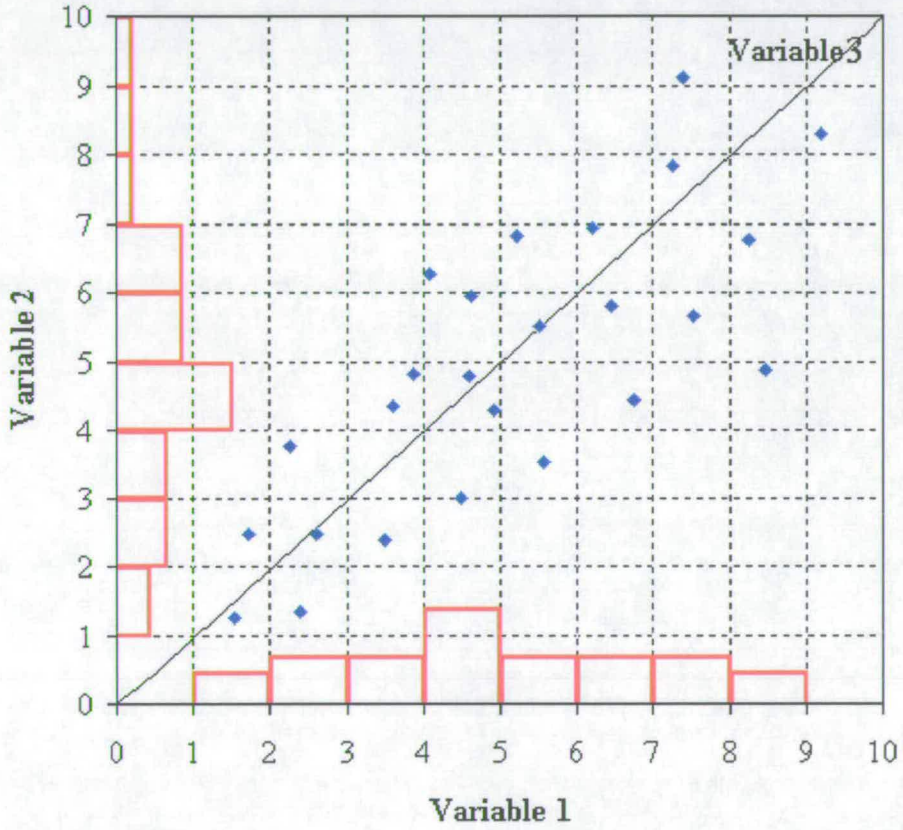


Figure 2.2 *Bivariate data set displaying distribution frequency and new axis to describe data.*

The new variable can be defined as the sum of the original variables (Equation 2.13). The actual values for a and b are arbitrary but are usually defined as the normalised linear combination coefficients so that the new variables scale to the same range as the original axes. The power of this technique lies in finding the linear combinations of variables which describe the maximum amount of variance possible *i.e.* allow the

reduction of the dimensionality of the data[16]. This technique is described in the following section.

$$Var3 = a.Var1 + b.Var2 \qquad \textbf{Equation 2.13}$$

2.1.7.2 Principal component analysis

Principal component analysis (PCA) is the most common of the techniques for combining variables[17]. It permits description of a data set using fewer variables than originally used. It may also reveal the variables or combination of variables that form some inherent structure within the data set. This structure may relate to some physio-chemical properties of the objects analysed[18]. In the case of HFC 134a it may reveal which by-products are the most important for defining different samples. The description of PCA that follows uses a simple two dimensional model as an example. The equations and methods used are of the same form regardless of the number of variables.

Consider a set of values for two variables which are used to describe a series of samples. The variance and covariance of these two variables can be calculated and represented in a variance/covariance matrix (Equation 2.14).

$$Cov_{X1,X2} = \begin{array}{c|cc} & X1 & X2 \\ X1 & Var_{X1} & Cov_{X2,X1} \\ X2 & Cov_{X1,X2} & Var_{X2} \end{array} \qquad \textbf{Equation 2.14}$$

These values can be displayed on a graph of variances of each variable (Figure 2.3). The values on each axis represent the variance within a variable and the length of each line represents the covariance between the two variables. If the data follow a multi-variate normal the two points marked lie on the boundary (or surface) of an ellipse whose centre is the origin (as stated in Section 2.1.6). The major axis of the ellipse is the eigenvector associated with the first principal component. This line represents the axis which describes the maximum amount of variation within the data. The second principal component is the minor axis of the ellipse. As the second axis is orthogonal to the first axis, the two axes are uncorrelated. Thus the new axes

onto which the original data are transformed are the eigenvectors of the ellipse derived from the variance/covariance diagram[19].

When the principal components are calculated for three variables the points of the variance/covariance diagram lie on the boundary of a 3-D ellipsoid, in which the third principal axis is again orthogonal to the first two. Objects defined by more than three variables have variance/covariance points lying on multidimensional elliptical shapes which cannot be visualised but do follow the same equations.

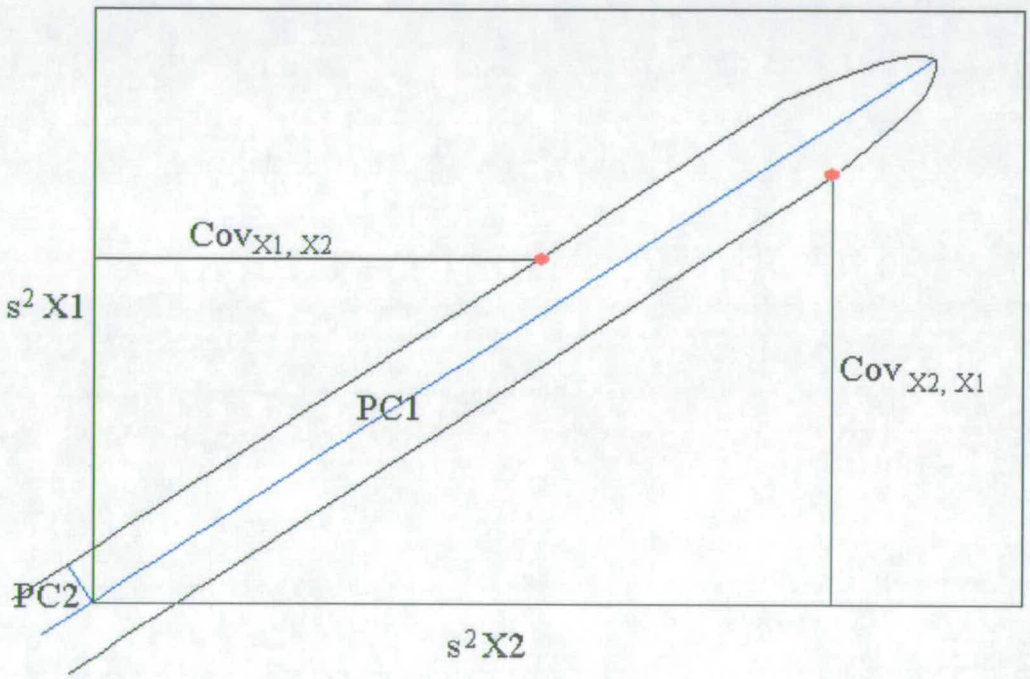


Figure 2.3: *Variance-covariance diagram for bivariate data showing PC1 and PC2 as the major and minor axes of an ellipse.*

The length of the eigenvector is directly proportional to the amount of variance that the axes describes. The calculation of the eigenvector is performed by a computer program where the values for the eigenvector describe the slope of the new axes (principal component). The original data are transposed onto the new axes using the values from the eigenvector which are termed 'the loadings'. The transposed values are termed the scores and are used to describe the data[20].

The variance described by each principal component can be displayed on a graph as a scree plot[21]. The scree plot shows the eigenvalue for each principal component (Figure 2.4, page 50). The total number of principal components needed to describe all the variance in the original data remains the same as the number of original variables. However, the majority of the variance is usually described in the first few principal components, negating the need to consider the remaining components.

2.1.7.3 Factor analysis

A closely related technique to principal component analysis is factor analysis. This is where the data described by m variables is reduced to fewer variables p (or factors) The new axes that these factors form are rotated in multidimensional space. The rotation is an attempt to assign some meaning to objects' positions on the factor axes. Factor analysis is performed by extracting the eigenvalues and eigenvectors from the variance/covariance or correlation matrix of a multivariate data set. The correlation matrix is formed by standardising the variance/covariance matrix.

There are five basic steps in a factor analysis.

1: Pre-processing of raw data matrix.

2: Calculation of the variance/covariance or correlation matrix.

3: Eigen analysis of the variance/covariance or correlation matrix.

4: Selecting the appropriate number of factors to describe the data.

5: Rotating the factor axes in order to impart meaning to position of objects on new axes.

Steps 1-3 are the same as in principal component analysis. Step 4 considers the number of factors required to describe the data. The calculations performed give a set of values showing how much of the original variation is described by each of the new factors. This can either be viewed graphically on a scree plot (Figure 2.4) or numerically as the cummulative % of original variation (Table 2.5).

There are objective mathematical functions which can be applied to decide the number of factors which best describe the data but usually it is possible to decide on viewing the cummulative % values. This allows the factors which describe residual variation *i.e.* noise, to be excluded from the interpretation of the data. In the example displayed factors 1, 2 and 3 describe the majority of variance in the original data. This allows factors 4 to 12 to be discounted from any further analysis[22]. Therefore in the example shown the dimensionality of the data set has been reduced from 12 variables to 3 factors.

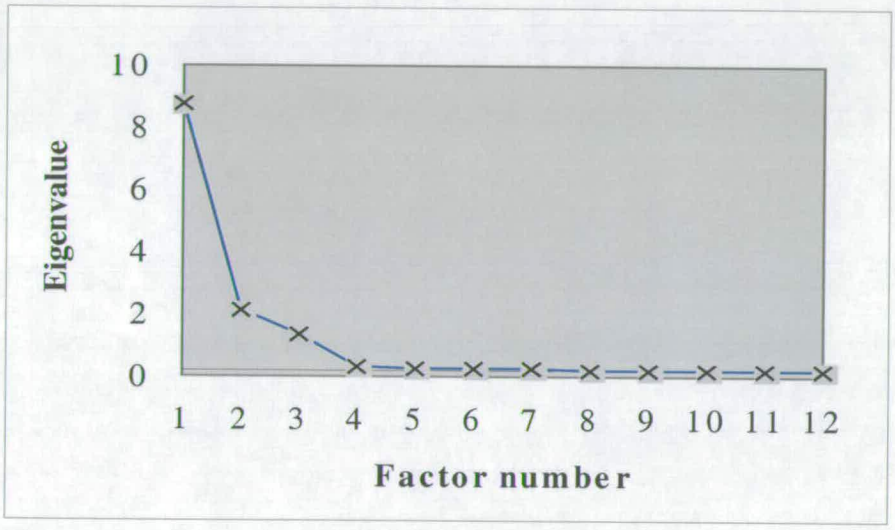


Figure 2.4 *Scree plot for data displayed in Table 2.5.*

Table 2.5 *Table describing variance distribution obtained from a factor analysis.*

Factor number	Eigenvalue	% Original variance	Cummulative variance
1	8.554	71.30	71.3
2	1.925	16.06	87.4
3	1.160	9.68	97.0
4	0.144	1.20	98.2
5	0.089	0.74	99.0
6	0.046	0.38	99.4
7	0.031	0.26	99.6
8	0.022	0.18	99.8
9	0.014	0.12	99.9
10	0.008	0.07	100.0
11	0.003	0.03	100.0
12	0.001	0.01	100.0

The data set can be considered in three dimensions representing the three factors used to describe the data. All other factors can be discounted from the analysis. The remaining three factor axes can be rotated in multidimensional space as they are no longer constrained by the need to be orthogonal to all other factors. As long as the three factors are orthogonal to each other they can be rotated. The most common type of rotation is *varimax* where the factor axes are rotated so that the objects they describe are either near the origin or the extremities of the factor axes. This will impart some meaning to the position of the objects in factor space[23].

2.2 Unsupervised pattern recognition

Classification of objects is a useful and often necessary process[24]. Whether it be the classification of elements in the periodic table or samples of HFC 134a into sites of production, treating single objects as parts of a group allows for shorter processing times. Although the human brain is excellent at recognising and classifying patterns and shapes it performs less well when presented with lists of numerical data. Consider Table 2.6 and Figure 2.5. Each displays the same set of data of six samples defined by five measured variables. The pattern seen in Figure 2.5 is obvious, Samples A, D and E form one group and Samples B, C and F form another group; however, the tabulated data are not so clear. Furthermore, with increased numbers of variables and samples to consider, the ability to visualise the data becomes difficult, if not impossible.

Table 2.6: *Data set for six samples defined by five variables.*

Variables	Sample A	Sample B	Sample C	Sample D	Sample E	Sample F
1	2.1	6.2	5.1	1.8	1.9	5.6
2	4.1	1.8	1.6	5.2	5.7	2.2
3	4.9	1.9	1.7	5.1	4.8	1.8
4	4.6	5.5	5.4	4.4	4.2	5.3
5	2.7	4.8	4.2	3.3	3.1	4.4

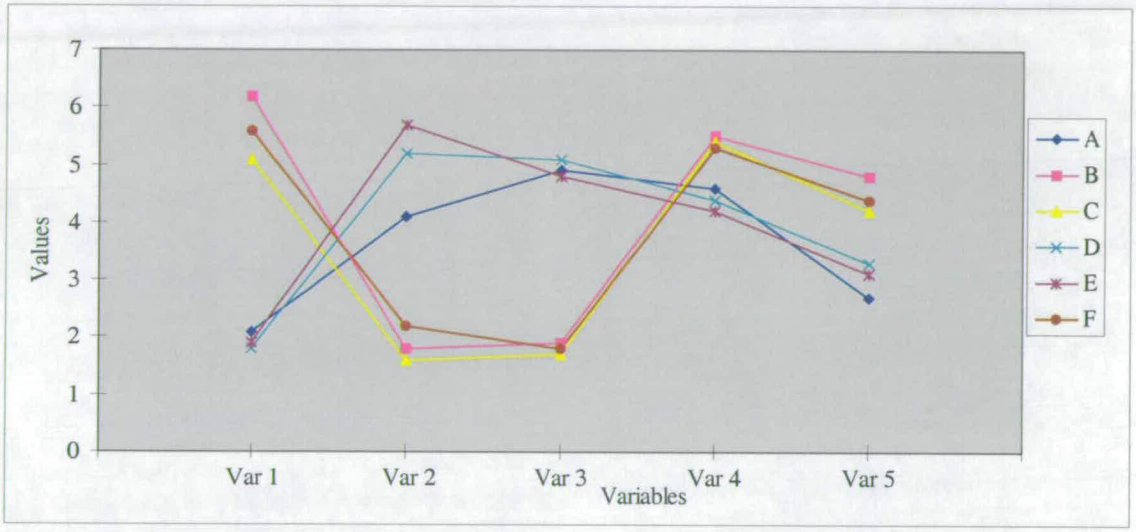


Figure 2.5 *Graphical representation of data set from Table 2.6.*

As modern analytical techniques are able to generate large quantities of data with many variables it is necessary to apply formal mathematical methods to highlight similarities and differences between samples. The problem can be stated as: given a number of samples, each defined by a set of measured variables, is it possible to derive a formal mathematical scheme for grouping these samples into classes such that samples within a class are similar, and are different from those in other classes? The number of classes and the characteristics defining each class are not known. The final statement is what discriminates between unsupervised pattern recognition and supervised pattern recognition. Supervised pattern recognition refers to the process of assigning new samples into known clusters depending on similarities and is discussed in detail in Section 2.3.

2.2.1 Measures of similarity and dissimilarity

Unsupervised clustering techniques generally start with a measure of the similarity or dissimilarity between the objects under analysis. The cluster membership and

number of clusters may depend on the metric used to characterise similarities/dissimilarities. Matrices of similarities between objects and distances between objects are complementary concepts although the distance measure is more often used.

2.2.1.1 Similarity measures

Similarity measures are normally defined using the correlation coefficient (as calculated in Section 2.2.5). Objects which show the highest correlations (values approaching 1 or -1) are considered similar and those which show low correlations (values approaching -0) are considered dissimilar.

2.2.1.2 Distance measures

Distance measures are the most commonly used method for calculating similarities between samples. The process works as follows: consider a set of samples defined by values obtained for measuring a number of different variables. Each sample is defined as a point plotted in multi-dimensional space where each dimension is defined by a variable. The similarity of two samples is the distance between their two points in multi-dimensional space[25].

There are numerous functions used to calculate the distance between two points but by far the most common is the Euclidean distance (Equation 2.15), where x_{ij} is the value of the j^{th} variable measured on the i^{th} sample. The data may be transformed before distance calculations are performed if the variables measured are in different units or have grossly differing means.

$$d_{AB} = \left[\sum_j (x_{1j} - x_{2j})^2 \right]^{1/2} \qquad \textbf{Equation 2.15}$$

2.2.2 Cluster definition

The general scheme to follow when performing unsupervised pattern recognition (cluster analysis) proceeds as follows. The data set comprising the original, or processed, data is transformed into a set of similarity or dissimilarity measures between the samples. The clustering process seeks to form a number of clusters so that the distance between similar samples within a cluster is minimised and distances between clusters is maximised. It is the concept of similarity between samples that allows for the diversity of techniques available in cluster analysis. *E.g.* consider Figure 2.6 which represents 62 objects defined by values for two variables.

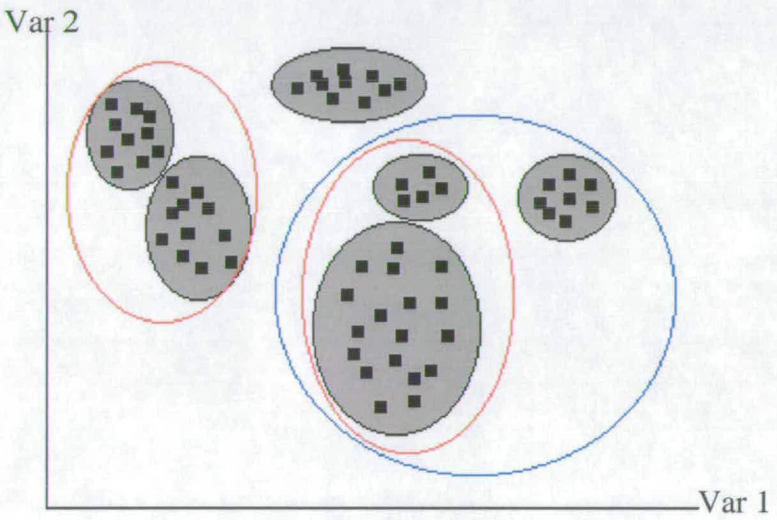


Figure 2.6 *Different cluster boundaries drawn for a 2-dimensional data set.*

Obviously there are a number of different interpretations of the clustering pattern which can be applied with no one answer being the 'correct' one. It is therefore necessary to carefully consider the methods and results of a cluster analysis, as with any analytical technique.

2.2.3 Hierarchical clustering

Hierarchical clustering can either be agglomerative, starting with all samples as individuals and linking them into one group[26], or divisive, starting with one cluster and continually splitting this group into its individual samples. The techniques employed in this thesis use the agglomerative process. The clustering technique

works by viewing each sample as a point in multidimensional space where each variable describes a dimension. The distance between the samples is calculated and the closest two samples linked into a cluster. This cluster is then viewed as a single sample and the next two closest samples are linked. This linking continues until all samples are grouped into one cluster. The overall process can be thought of in four steps:

Step 1- the distance between all the samples is calculated in multidimensional space

Step 2- the two closest samples are linked and considered as a single point in multidimensional space

Step 3- the new distance between samples is calculated taking into account the new cluster formed in Step 2

Step 4- the process returns to Step 2 until all the samples have be linked into one cluster

Different methods of hierarchical clustering are produced by using different calculations for the distance measures between the samples and the method used to describe a new sample after two samples have been linked. The more common calculations to describe the distances between a sample and a cluster are as follows:

Nearest neighbour

The distance between a sample and a cluster is given by calculating the distance between the sample and the closest member of the cluster to which the sample is being linked, as shown in the Figure 2.7.

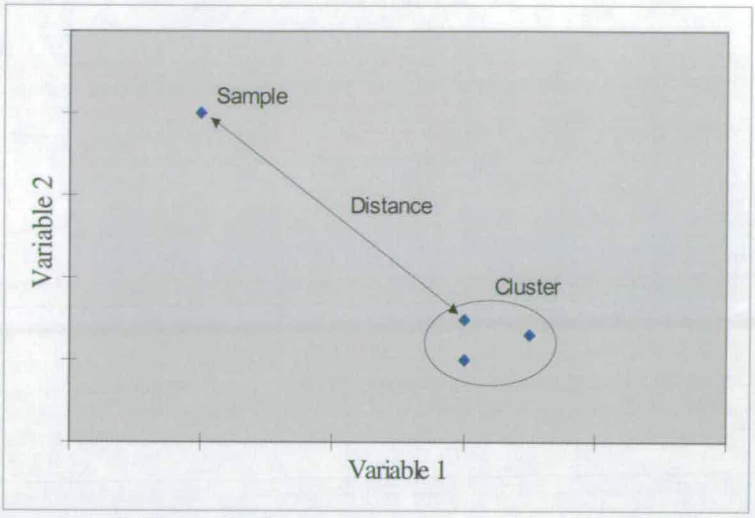


Figure 2.7 *Nearest neighbour linkage*

Furthest Neighbour

The distance between a sample and a cluster is given by calculating the distance between the sample and the furthest member of that cluster from that sample as shown in Figure 2.8.

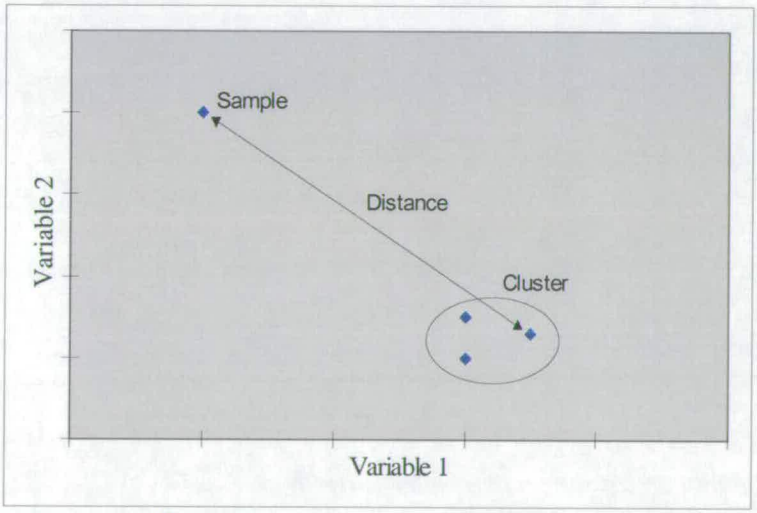


Figure 2.8 *Furthest neighbour linkage*

Average linkage method

When a cluster of samples is linked to another sample the distances between the new sample and each point in the cluster is considered. In the example shown (Figure 2.9) the three distance measures are taken and averaged to find the distance measure. If two clusters are to be joined the distance measures between all the samples in each cluster are taken and averaged.

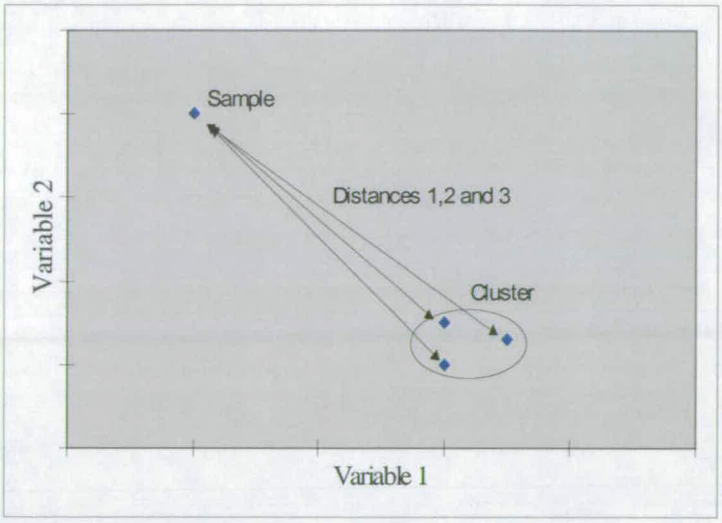


Figure 2.9 *Average linkage method*

Whichever method is used the results of hierarchical clustering are usually presented in the format of a dendrogram. The dendrogram is constructed by plotting the distance measures against the samples linked (Figure 2.10, page 58). This forms a tree-like structure which can be easily interpreted in terms of the different samples and their cluster memberships.

2.2.3.1 Example of hierarchical clustering

A simple two dimensional data set will be used as an example of the steps taken and results obtained in hierarchical clustering. The bi-variate data set is shown in Table 2.7 and consists of seven samples described by two variables (Var 1 and Var2). The first stage of the analysis involves the inspection of the data to decide if any pre-processing is required. In this case the values are in the same range and neither variable will dominate the distance measures.

Table 2.7 *Bivariate data set for seven samples*

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Sample 7
Var 1	4.0	4.0	4.5	3.0	2.7	3.5	1.0
Var 2	1.0	1.5	1.3	3.0	2.8	3.0	4.6

The distances between all the samples is calculated using the Euclidean distance measure. The distance matrix is displayed in Table 2.8. The matrix is symmetrical

about the diagonal with values along the diagonal of zero, as sample 1 to sample 1 distance is zero.

Table 2.8 *Euclidean distance matrix for bivariate data from Table 2.7.*

	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Sample 7
Sample 1	0.50	0.58	2.24	2.22	2.06	4.69
Sample 2		0.54	1.80	1.84	1.58	4.31
Sample 3			2.27	2.34	1.97	4.81
Sample 4				**0.36**	0.50	2.56
Sample 5					0.82	2.47
Sample 6						2.97

Table 2.9 shows the hierarchical clustering scheme for the distance matrix in Table 2.8 using the nearest neighbour methodology. The use of the word Cluster in Table 2.9 refers both to single samples and groups of samples where the cluster is named after the lowest sample number it contains. Stage 1 in the process locates the two closest samples, 4 and 5 with a distance measure of 0.36 (bold in Table 2.8), and joins them as a cluster. This cluster next appears in Stage 2. Stage 2 locates the next closest clusters, sample 6 and cluster 4. The distance between sample 6 and cluster 4 is taken as 0.50, the distance between 5 and 6 the nearest neighbours.

The linkage continues until finally cluster 1 (representing samples 1,2,3,4,5 and 6) is linked with sample 7; the distance value being taken between samples 7 and 5, the nearest neighbours. The data in Table 2.9 can be plotted onto a dendrogram (Figure 2.10) to allow easier interpretation of the groups formed in the cluster analysis. The distance measures are linearly rescaled to values between 0 and 25. From the dendrogram it can be assumed that the data fall into three groups (1,2,3), (4,5,6) and (7).

Table 2.9 *Hierarchical clustering for bivariate data using nearest neighbour linkage*

	Clusters combined			Stage cluster first appears		
Stage	Cluster 1	Cluster 2	Distance	Cluster 1	Cluster 2	Next Stage
1	4	5	0.36	0	0	2
2	4	6	0.50	1	0	5
3	1	2	0.50	0	0	4
4	1	3	0.54	3	0	5
5	1	4	1.58	4	2	6
6	1	7	2.47	5	0	0

Close consideration should be taken when applying hierarchical clustering as the clusters formed distort the data somewhat. In the example given the nearest neighbour method reduces the apparent distance between samples *i.e.* it appears that cluster 1 and its components are a distance of 2.47 from sample 7, when in fact samples 1,2 and 3 are all > 4.30 distance from 7.

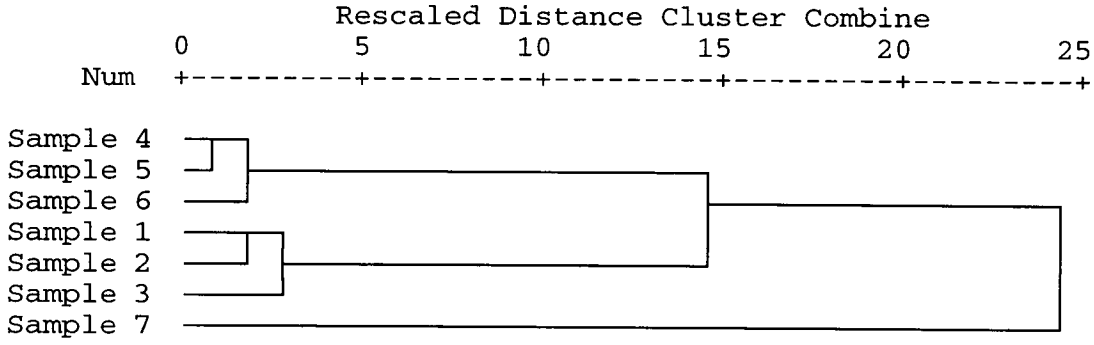


Figure 2.10 *Dendrogram formed from the hierarchical clustering of bi-variate data set from Table 2.1 using Euclidean measure and nearest neighbour linkage*

2.2.4 K-means algorithm

The K-means algorithm is a clustering technique which attempts to place *n* objects described by *m* variables into K clusters. Although this is an unsupervised pattern recognition technique, a value for the number of clusters required to describe the samples is entered before the analysis proceeds. This does not mean that a *priori* knowledge of the data is required as a number of separate analyses can be performed using various K numbers to provide the best fit for the data[27].

The technique can be described in four steps:

Step 1 Given a set number of clusters, K, and the initial membership of those clusters, calculate the total distance between the members and the cluster centres, the partition coefficient.

Step 2 Move one object to each of the different clusters and recalculate the partition coefficient. If the partition coefficient decreases place the object in that cluster and recalculate the cluster centres.

Step 3 Repeat step 2 for all objects.

Step 4 If no object has been moved then stop, or return to step 2.

As the technique is one of optimisation it is possible to obtain more than one cluster pattern depending on the initial cluster membership selected. The technique may find only a local minimum and therefore it is necessary to perform the analysis a number of times under different conditions to check the results obtained. Conditions to change could include initial cluster assignment of objects or the distance measure used.

2.2.4.1 Example of K-means clustering

A simple two dimensional data set can be used to show the calculation steps performed in a K-means cluster analysis. The data in Table 2.10 show the values for five samples (A, B, C, D and E) described by two variables (Var1 and Var 2).

Table 2.10 *Bivariate data set for five samples.*

Sample	A	B	C	D	E
Var1	1	1	4	5	4
Var 2	1	2	2	2	1

The initial cluster assignment is based on the value for Var 2, following the rules in Equation 2.16 which states: if the value for Var 2 lies above the mean of all the values then assign to Group B; if the value for Var 2 lies below the mean assign to Cluster B.

$$\left(Var2_i \rangle \overline{Var2}\right) \therefore Object_i = ClusterB$$
$$\left(Var2_i \langle \overline{Var2}\right) \therefore Object_i = ClusterA \qquad \textbf{Equation 2.16}$$

The initial cluster membership (Cluster 1) will therefore be Cluster A (A and E) and Cluster B (B, C and D). The initial cluster centres are calculated as the average points in space depending on the samples they contain.

$\therefore$ $CC_A = [(Var1_A+Var1_E)/2], [(Var2_A+Var2_E)/2] = \mathbf{(2.5, 1)}$

$CC_B = [(Var1_B+Var1_C+Var1_D)/3], [(Var2_B+Var2_C+Var2_D)/3] = \mathbf{(4.3, 1.3)}$

The distance between each object and the cluster centre it has been assigned to is calculated using the Euclidean distance measure (Equation 2.14, Section 2.3.1.2). These distance measures are used to calculate the partition coefficient (*e*), where the partition coefficient is the sum of all the distances.

$e = (A \rightarrow CC_A)^2 + (B \rightarrow CC_B)^2 + (C \rightarrow CC_B)^2 + (D \rightarrow CC_B)^2 + (E \rightarrow CC_A)^2$

$e = (2.5\text{-}1)^2 + (1\text{-}1)^2 + (4.3\text{-}1)^2 + (1.3\text{-}2)^2 + (4.3\text{-}4)^2 + (1.3\text{-}2)^2 + (4.3\text{-}5)^2 + (1.3\text{-}2)^2 + (2.5\text{-}1)^2 + (1\text{-}1)^2 =$ **17.44**

The value of 17.44 is a measure of how well the objects have been grouped. The objects are now moved between clusters to try to minimise the partition coefficient. In the example object E is moved to cluster B. The partition coefficient is again calculated.

$e = (A \rightarrow CC_A)^2 + (B \rightarrow CC_B)^2 + (C \rightarrow CC_B)^2 + (D \rightarrow CC_B)^2 + (E \rightarrow CC_B)^2$

$e = (2.5\text{-}1)^2 + (1\text{-}1)^2 + (4.3\text{-}1)^2 + (1.3\text{-}2)^2 + (4.3\text{-}4)^2 + (1.3\text{-}2)^2 + (4.3\text{-}5)^2 + (1.3\text{-}2)^2 + (4.3\text{-}4)^2 + (1.3\text{-}1)^2 =$ **15.37**

Object E is placed into cluster B as this reduces the partition coefficient and therefore is a better clustering model. The cluster centres are recalculated with the new cluster memberships where Cluster A (A) and Cluster B (B, C, D and E).

$CC_A = (Var1_A), (Var2_A) =$ **(1,1)**

$CC_B = (Var1_B+Var1_C+Var1_D+Var1_E)/4, (Var2_B+Var2_C+Var2_D+Var2_E/4] =$ **(3.5,1.8)**

The partition coefficient is calculated for this cluster model.

$e = (A \rightarrow CC_A)^2 + (B \rightarrow CC_B)^2 + (C \rightarrow CC_B)^2 + (D \rightarrow CC_B)^2 + (E \rightarrow CC_B)^2$

$e = (1\text{-}1)^2 + (3.5\text{-}1)^2 + (3.5\text{-}4)^2 + (3.5\text{-}4)^2 + (3.5\text{-}5)^2 + (1\text{-}1)^2 + (1.8\text{-}2)^2 + (1.8\text{-}2)^2 + (1.8\text{-}2)^2 + (1.8\text{-}1)^2 =$ **9.76**

Object B is now moved into Cluster A and the partition coefficient is recalculated.

$e = (A \rightarrow CC_A)^2 + (B \rightarrow CC_A)^2 + (C \rightarrow CC_B)^2 + (D \rightarrow CC_B)^2 + (E \rightarrow CC_B)^2$

$e = (1\text{-}1)^2 + (1\text{-}1)^2 + (3.5\text{-}4)^2 + (3.5\text{-}4)^2 + (3.5\text{-}5)^2 + (1\text{-}1)^2 + (1\text{-}2)^2 + (1.8\text{-}2)^2 + (1.8\text{-}2)^2 + (1.8\text{-}1)^2 =$ **4.47**

The partition coefficient is again reduced showing that this is a better description of cluster membership. Object B is assigned to Cluster A and the partition coefficient recalculated. The process continues until a minimum value is obtained. By plotting the values in Table 2.10 on a graph where each axis is a variable (Figure 2.11) it can be seen that cluster model Cluster A(A and B) and Cluster B (C, D and E) best describes the data for K=2.

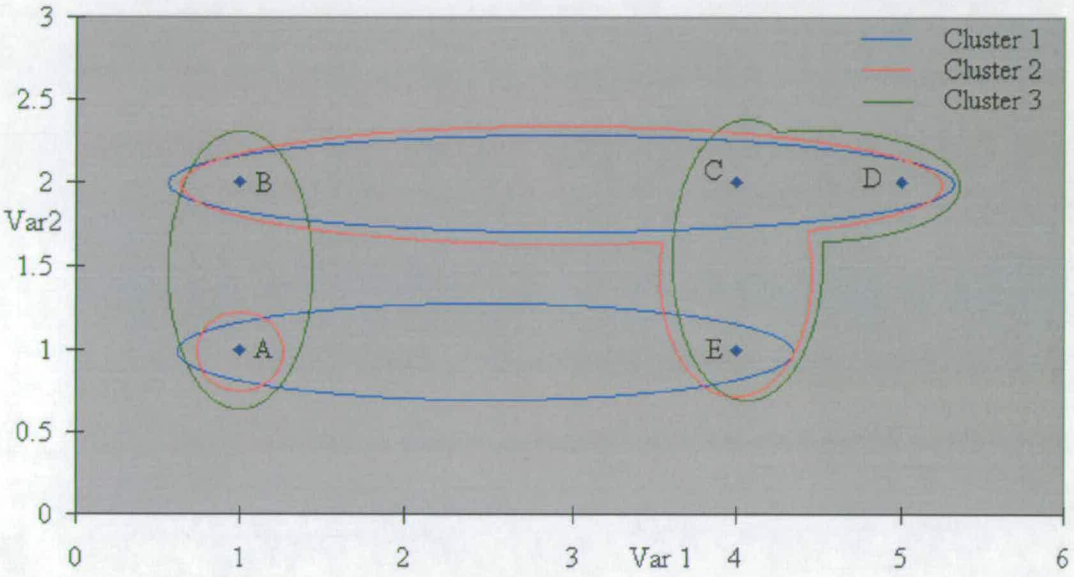


Figure 2.11 *Cluster patterns formed by K-means cluster for bivariate data.*

2.3 Supervised Pattern recognition

Supervised pattern recognition deals with the following problem; given a set of data describing samples of known different parent populations is it possible to define a function that will allow an unknown sample to be classified into one of these parent groups?[28] In the context of this work an example would involve analysing a number of HFC 134a samples of known origin and forming a discriminant function that allows new samples of unknown origin to be identified. There are a number of different techniques used to form discriminant functions but they all require a training data set. That is a data set which adequately describes the different parent populations to be discriminated. The different techniques all follow the same general scheme:

Step 1 Investigate the data of known parent populations to form a

discriminant function.

Step 2 Test the discriminant function using test data.

Step 3 Apply the discriminant function to unknown data.

The technique applied to investigate the data is closely related to the form of the data under study. If the parent populations of the data are known to follow normal distributions then parametric techniques such as statistical discriminant analysis can be performed. If the distribution of the data is unknown or not normally distributed then non-parametric techniques such as K-nearest neighbour are applied. Other techniques such as artificial neural networks can be applied to distinguish samples although these are not considered in this work.

Testing of the discriminant function can be performed by using new data of known parent population, data used in the formation of the discriminant function or the 'leave one out' method. The best method is to use new data of known parent population to test the discriminant function. This method, however, does have a number of drawbacks. The larger the parent data set the better the discriminant function will be. Therefore, using data of known parent population in testing is somewhat wasteful especially if data sets are small *i.e.* this test data could be used to form a better discriminant function. The second method is to test the discriminant function using the data from the parent populations, although this will have a lowered biased error value. The final method is to form the discriminant function using all the data apart from one sample. This sample is used to test the function. The discriminant function is then recalculated leaving out another sample which is used to test the function. This procedure is used for all the samples making up the parent population. The main drawback with this method is that the true error calculations relating to all the samples being used are never made.

2.3.1 Statistical discriminant functions

Statistical discriminant analysis is a widely used parametric method for assigning unknown objects into parent groups. A simple bivariate data set will be used to

describe the method. Table 2.11 contains the data representing two parent populations described by two variables.

Table 2.11 *Bivariate data set for samples belonging to two clusters*

Data representing population 1			Data representing population 2		
Sample	Var 1	Var 2	Sample	Var 1	Var 2
1	3.2	2.0	9	5.7	3.5
2	2.1	2.2	10	5.2	4.3
3	2.6	2.3	11	5.9	4.9
4	2.4	2.9	12	4.3	5.4
5	3.1	3.4	13	5.2	5.7
6	3.9	4.0	14	6.0	6.3
7	3.7	3.1	15	4.7	7.0
8	3.9	3.5	16	3.7	7.3
Mean	3.12	2.92		5.09	5.55
Variance	0.47	0.51		0.66	1.71

The calculation of the descriptive statistics shows that samples belonging to population 2 have generally higher values than those in population 1. By approaching the problem using a single variate analysis, a discriminant function could be based on the mean values for Var 1 or for Var 2. An example of a discriminant rule based on Var 1 would be as follows; a sample is assigned to population 1 if its Var 1 value is smaller than the mean of the two parent populations (Equation 2.17) and to population 2 if the value is greater than the mean (Equation 2.18).

Assign new sample to population 1 if

$$Var1 < (\bar{x}_{Cluster1} + \bar{x}_{Cluster2})/2 \qquad \textbf{Equation 2.17}$$

Assign new sample to population 2 if

$$Var1 > (\bar{x}_{Cluster1} + \bar{x}_{Cluster2})/2 \qquad \textbf{Equation 2.18}$$

The results of applying these rules based on Var 1 and Var 2 in isolation are displayed in Table 2.12, which is described as a confusion matrix. This displays the results from testing the rule using the parent populations. E_{ij} is the number of samples from population i that have been classified as belonging to population j. M_{ia} is the number of samples actually in population i and M_{ic} is the number of samples the rule has placed into population i. From the confusion matrix the mean rule used for either variable alone leads to misclassification of samples.

By plotting the data points on a bivariate axes (Figure 2.12) it is clear that the data sets fall into two distinct clusters and a line could be inserted as the discriminant function. It is the positioning of this diagonal line which is the key to a successful discriminant function. The line of best fit is calculated by considering probability and Bayes theorem.

Table 2.12 *Confusion matrix used to classify results of discriminant analysis using Var 1 and Var2.*

		Actual membership		
		Cluster1	Cluster 2	
Predicted	Cluster 1	E_{11}	E_{21}	M_{1c}
membership	Cluster 2	E_{12}	E_{22}	M_{2c}
		M_{1a}	M_{2a}	

Discriminant rule based on Variable 1

		Actual membership		
		Cluster 1	Cluster 2	
Predicted	Cluster 1	8	1	9
membership	Cluster 2	0	7	7
		8	8	

Discriminant rule based on Variable 2

		Actual membership		
		Cluster 1	Cluster 2	
Predicted	Cluster 1	8	1	9
membership	Cluster 2	0	7	7
		8	8	

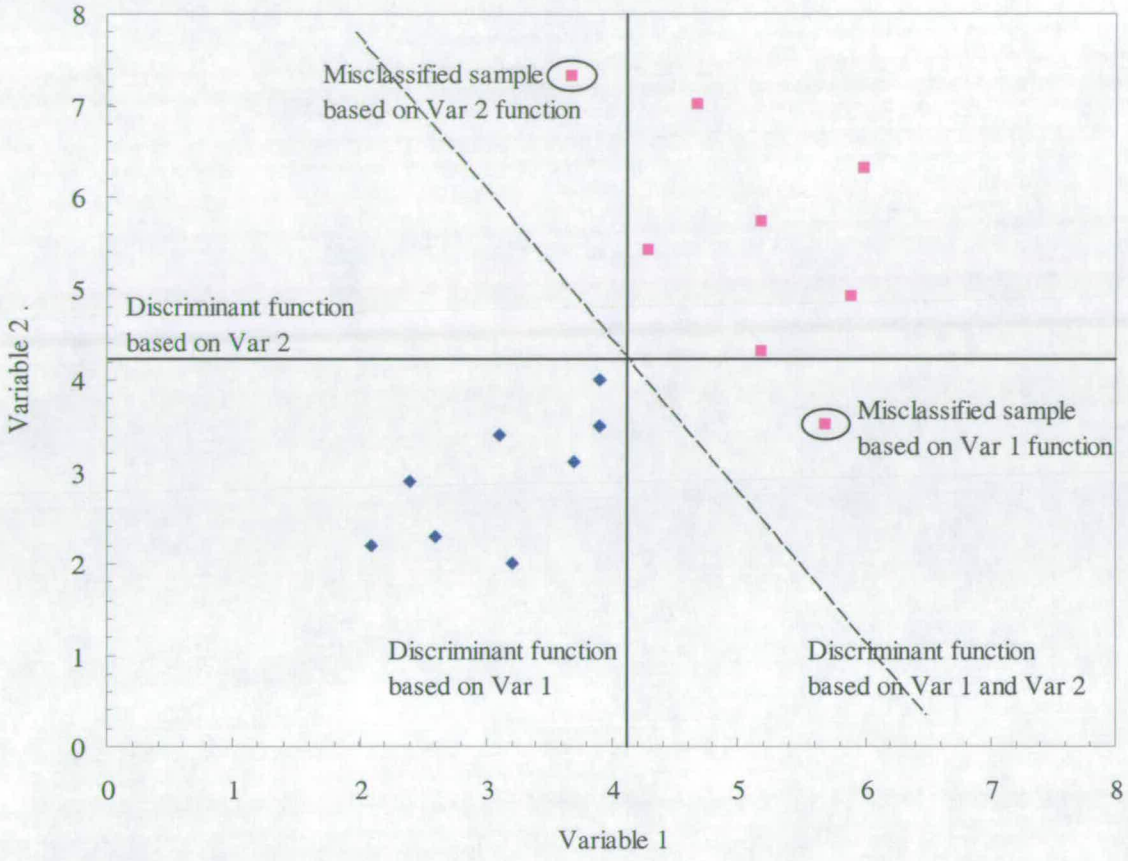


Figure 2.12 *Data set from Table 2.1 projected onto bivariate axes.*

2.3.1.1 Bayes theorem

Bayes rule simply states that 'any unclassified object should be assigned to the cluster to which it has the highest conditional probability'[29]. The term conditional probability can be explained in terms of rolling a six-sided die. The probability of rolling a six is 1/6 and the probability of rolling two dice and both showing a 6 is 1/36 (1/6*1/6). If, however, one of the dice is already showing a six then the probability of rolling a second six is 1/6. In the same way the probability of rolling two sixes if one die does not show a six is 0. This means that the probability of an outcome is altered either in favour or against due to *priori* knowledge. Applying this rule to the data in Table 2.11 where the priori knowledge is the values for variables 1 and 2 gives Equation 2.18.

Unknown sample is placed in Cluster 1 if

$$P_{(G(1)|Var1,Var2|)} \rangle P_{(G(2)|Var1,Var2|)} \qquad \textbf{Equation 2.19}$$

Where $P_{(G(1)|Var1,Var2|)}$ is the probability that the unknown belongs in Cluster 1 based on priori knowledge of the values for Var 1 and Var2
$P_{(G(2)|Var1,Var2|)}$ is the probability that the unknown belongs in Cluster 2 based on priori knowledge of the values for Var 1 and Var2

The conditional probability values expressed in Equation 2.19 are calculated from all the possible combinations of values for the variables describing a cluster being studied. In practice this is not possible and so an estimation of the conditional probability is calculated according to Bayes Theorem. The equation for Bayes Theorem still contains a value for the conditional probability of a sample belonging to any one cluster. However, as the variables contributing to the definition of the samples follow a multivariate normal distribution (as this is a parametric technique) the equation can be expressed in a quadratic form, of which the solution forms the discriminate function. In the case of bivariate data the discriminant function is a curved line, in the case of trivariate data it is a curved plane and in multivariate data it has multidimensional boundaries.

2.3.2 K-Nearest neighbour

One of the most widely used non-parametric techniques is that of the K-nearest neighbour calculation (K-NN). This method views each sample in the parent population in multi-dimensional space where each dimension represents a variable. An unclassified sample is placed into this space and the Euclidean distances between it and all the other samples are calculated. The new sample is placed into the cluster with which it shares the majority of its nearest neighbours.

Figure 2.13 shows the distribution of two parent populations (Cluster 1 and Cluster 2) described by two variables. An object of unknown Cluster origin is plotted onto the graph. The K-NN analysis is performed on the data set and the five nearest neighbours are selected. In the case of bivariate data this selection is represented as a circle of radius r where the magnitude of r defines the number of neighbours to be considered. The cluster membership of the nearest neighbours is checked and the

unknown is assigned to the cluster with most of its nearest neighbours. In the example shown it is assigned to Cluster 1.

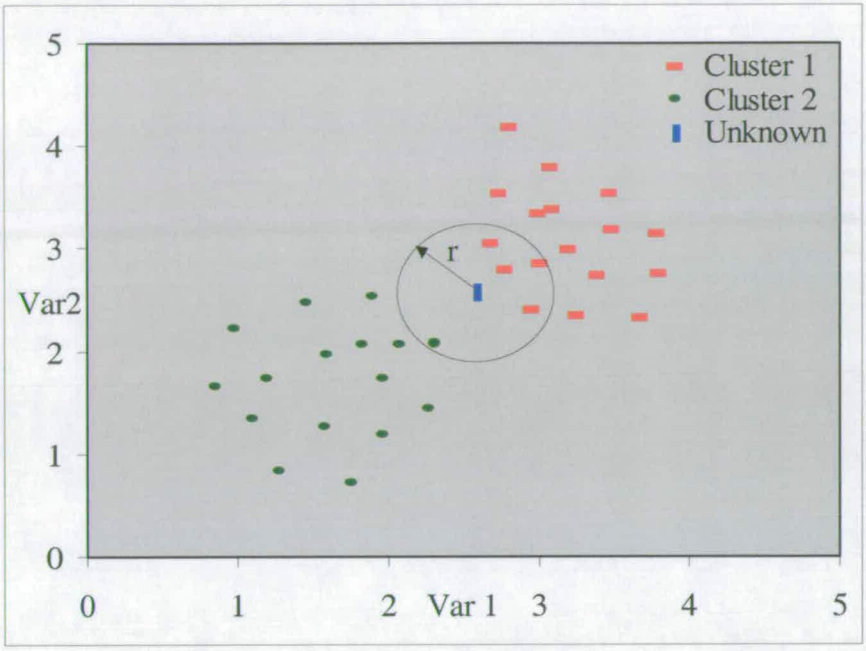


Figure 2.13 *Classification circle for an unknown object using K-NN for five nearest neighbours.*

2.4 Application of chemometrics to analysis of HFC 134a samples

As detailed earlier, the data obtained for the analyses of HFC 134a samples will consist of an *mxn* matrix of values. The rows (*n*) represent the different samples of HFC 134a analysed, the columns (*m*) represent the different variables (synthesis by-products present) and the values in the matrix are the measured responses of the by-products to the detector. The large number of by-products present in the different samples requires the use of multivariate analysis techniques[30].

The aim of these analyses is to classify different samples into different groups and to link similar samples together. Once this has been achieved the results will be investigated to ascertain which by-products are important in the classification of samples and whether any information on the routes of synthesis can be obtained. The classified samples will then be used to form a discriminant function with which new unknown samples will be identified.

The data from the GC-FID and GC-ECD analyses will be studied individually using the same chemometric techniques. This will give an indication as to which, if either, experimental analysis set up provides the best characterisation data.

The chemometric techniques will be applied to three forms of the data; the original untreated results from the analyses, termed the raw data; a data set which has been transformed to remove absolute magnitude effects of the different variables, termed the normalised data; and the results from linear combinations of the original data using factor analysis, termed the factor data.

The three data sets will be subjected to unsupervised cluster analysis using the K-means algorithm and Hierarchical clustering. The cluster information obtained from these analyses will then be used to form discriminant functions with which new samples of known and unknown origin will be classified.

The final result of the testing of the chemometric techniques will lead to the formation of a set protocol representing the best method for the classification and identification of different samples of HFC 134a (Figure 2.14).

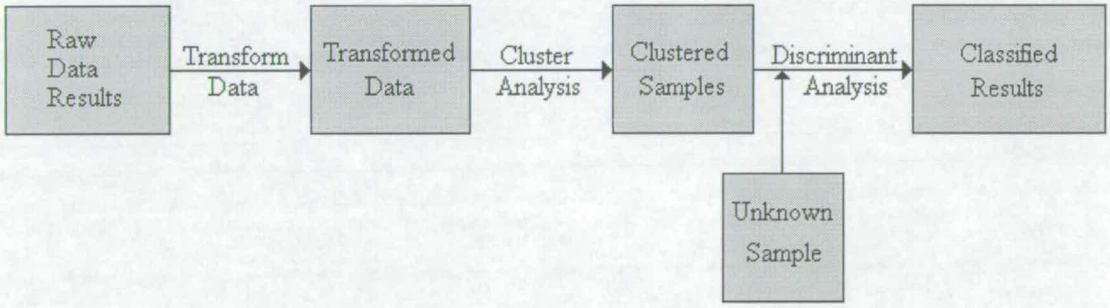


Figure 2.14 *Chemometric classification scheme for HFC 134a samples.*

References

[1] Shoenfeld, P.S. and Devoe, J.R., *Anal. Chem.*, **44**, 497R, (1976)

[2] Brown, S.D., Sum S.T., Despagne, K. and Lavine, B.K., *Anal. Chem.*, **68**, 21R-61R, (1998)

[3] Fredericks, P.M., Lee, J.B., Osborn, P.R. and Swinkels, D.A.J., *Appl. Spec.*, **39**, N2, 303-310, ()

[4] Adams, M.J., *Chemometrics in Analytical Spectroscopy*, Royal Society of Chemistry, Cambridge, (1995)

[5] Dyson, N., **Chromatographic Integration Methods**, p41-45, Royal Society of Chemistry, Cambridge, (1992)

[6] Mood, A.M., Graybill, F.A. and Boes, D.C., *Introduction to the Theory of Statistics*, p107-111, 3rd ed., McGraw-Hill International, London, (1974)

[7] Zak, R., *Basic Facts for Basic Science*, p36, Raven Press, New York, (1990)

[8] Beyer, W.H., *Standard Mathematical Tables*, 27th ed., CRC Press, Florida, p546, (1996)

[9] *GraphPad Prism™ User's Guide*, San Diego, p275 (1995)

[10] Morgan, E., *Chemometrics: Experimental Design*, John Wiley and Sons, Chichester, England, p20-25, (1995)

[11] Manly, B.F.J., *Multivariate Statistical Analysis: A Primer*, Chapman and Hall, London, UK, (1991)

[12] Buckland, W.R. and Kendall, M.G., *A Dictionary of Statistical Terms*, 4th ed., Longman Group Ltd, London, p46, (1982)

[13] Adams, M.J., *Chemometrics in Analytical Spectroscopy*, Royal Society of Chemistry, Cambridge, p21-22, (1995)

[14] Pearson, K., *Philosophical Magazine*, **6**, 2, 559-572, (1901)

[15] Stout, S.A., *Journal of Analytical and Applied Pyrolysis*, **18**, 277-292, (1991)

[16] Stout, S.A., *Journal of Analytical and Applied Pyrolysis*, **18**, 277-292, (1991)

[17] Brown, S.D., Sum, S.T., Despagne, F. and Lavine, B.K., *Anal. Chem.*, **68**, 21R-61R, (1996)

[18] Hellmuth, W.W. and Crawford, N.R., *Fuel*, **69**, 443-447, (1990)

[19] Wold, S., Esbensen, K. and Geladi, P., *Chemometrics and Intelligent Laboratory Systems*, **2**, 37-52, (1987)

[20] Aries, R.E., Lidiard, D.P. and Spragg, *Chem. Br.*, **27** p821, (1991)

[21] Adams, M.J., *Chemometrics in Analytical Spectroscopy*, Royal Society of Chemistry, Cambridge, p75-76, (1995)

[22] Hopke, P.K., Alpert, D.J. and Roscoe, B.A., *Comput. Che.*, **7**, p149, (1983)

[23] Flury, B. and Riedwye, H., *Multivariate Statistics: A Practical Approach*, Chapman and Hall, Lodon, (1984)

[24] Adams, M.J., *Chemometrics in Analytical Spectroscopy*, Royal Society of Chemistry, Cambridge, p92, (1995)

[25] Manly, B.F.J., *Multivariate Statistical Analysis: A Primer*, Chapman and Hall, London, UK, (1991)

[26] Alsberg, B.K., *Anal. Chem.*, **71**, 15, p3092-3100, (1999)

[27] Hertigan, J., *Clustering Algorithms*, John Wiley and Sons, Chichester, England, (1975)

[28] Everitt, B., *Cluster Analysis*, 2nd ed., Heinemann Educational, London, (1980)

[29] Mood, A.M., Graybill, F.A. and Boes, D.C., *Introduction to the Theory of Statistics*, p107-111, 3rd ed., McGraw-Hill International, London, (1974)

[30] Adams, M.J., *Chemometrics in Analytical Spectroscopy*, Royal Society of Chemistry, Cambridge, p22-26, (1995)

Chapter 3

Testing and Optimisation of Experimental Conditions

3.0 Introduction to chapter

This chapter describes the experimental equipment and methods used in the characterisation of the HFC-134a samples. Each section gives a brief description of the equipment or method to be used followed by the results of testing performed in order to validate that step in the analysis of HFC 134a. Validation of each step was performed in a sequential order of usage, *i.e.* sample dilution was validated before the injector system. The experimental techniques were rigorously tested to produce the best data possible. That is, any differences in the profiles of HFC 134a due to experimental errors or inconsistencies will have been minimised.

3.1 Standards and samples

To perform the characterisation work on HFC 134a a number of samples of other halocarbons were obtained. Ideally, samples of all the potential by-products of synthesis of HFC 134a are required. However, because of constraints on availability and expense, a smaller representative group of halocarbons were obtained and used.

The physical properties of most of the halocarbons, including HFC 134a, means that they are gases at room temperature and pressure[1]. Even the compounds with the higher boiling points, 27.1°C for HCFC 123, are volatile enough to be treated as gases. The following sections detail the acquisition of samples and the techniques involved in the handling and dilution of these samples.

3.1.1 Acquisition of standards and samples

Methane standards were purchased from Aldrich Chemical Company (Milwaukee, USA) and were 99.0% pure. Halocarbon standards were obtained directly from Fluorochem Limited (Old Glossop, UK) or from Fluorochem via ICI Chemicals & Polymers, Runcorn and were >99% pure. Samples of HFC-134a were obtained directly from ICI C&P, Runcorn or from The Chemical Engineering Department (Edinburgh University) courtesy of Colin Pritchard. All gas samples were obtained in either pressurised cylinders or in Teflon gas sampling bags (Alltech, USA).

3.1.2 Sample handling

Transfer of gas samples from cylinders was performed by direct connection of the cylinder via 1/16th copper or stainless steel tubing. The tubing was flushed with the sample gas from the cylinder before connection. Specific volumes of gas were transferred using a 5ml gas tight syringe (BSGE, Australia) with an on/off tap (Figure 3.1). All injections of samples were performed through rubber septa (Alltech, USA). These septa were replaced at regular periods of time to prevent gas leakage through wear[2]. All sample dilutions and mixes were stored in either Teflon gas sample bags (Alltech, USA) or glass dilution bulbs (designed and prepared in house).

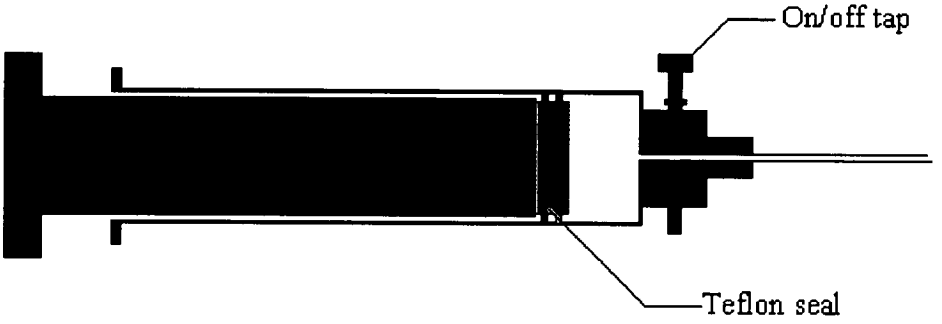


Figure 3.1 *Gas tight syringe with on/off tap*

3.1.3 Sample dilution and mixing

A glass vacuum vessel was designed and built in-house and used for the dilution and mixing of standards and samples. The design of the equipment is shown in Figure 3.2. Youngs taps (J Youngs Scientific, UK) are used to open and isolate different areas of the vessel to allow filling and evacuation of gas as required. A Speedivac ED75n high vacuum rotary pump (Edwards, UK), tested to $2x10^{-2}$ atm, was used to evacuate the system. The pump was connected to port 1 of the vessel using thick walled rubber tubing fastened with a screw clip. Helium used as dilution gas was taken, via a 1/16" Swagelock t-piece union, from the GC carrier gas line. Dilution gas pressure was controlled using a Porter pressure regulator and monitored by a Porter bourdon pressure gauge (0-30psi). The dilution gas line was connected to port 2 of the vacuum vessel using a Cajon Ultratorr connector (Cajon, UK).

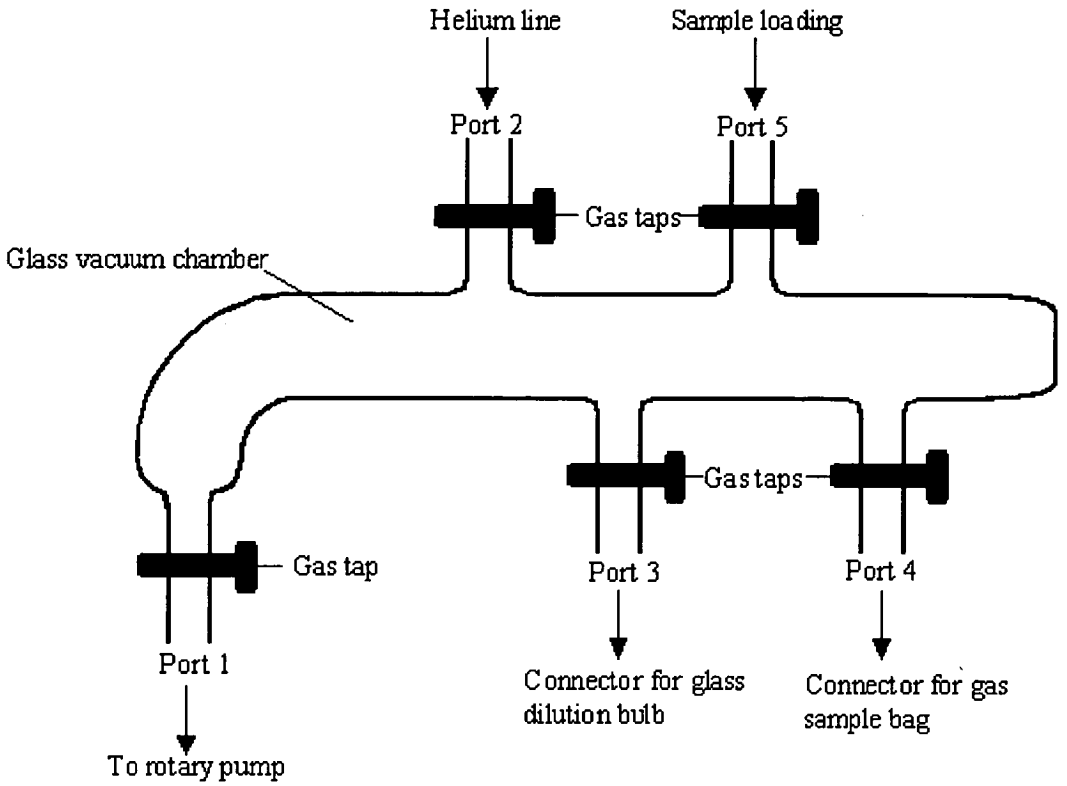


Figure 3.2 *Schematic of vacuum system used for gas handling*

Four glass dilution bulbs were designed and built in house (Figure 3.2). The four bulbs were labelled 1-4 and had volumes of 1074, 1086, 1078 and 1080ml respectively. Each bulb could be connected to port 3 of the vacuum vessel by means of a Youngs ball and cup joint sealed with a rubber O-ring and held in place using a Rotulex 19/9 spring clip (Pyrex, France). Teflon gas sample bags were connected to the vacuum vessel using thick walled rubber tubing via port 4. Port 5 of the vacuum vessel was fitted to allow direct loading of samples from gas cylinders.

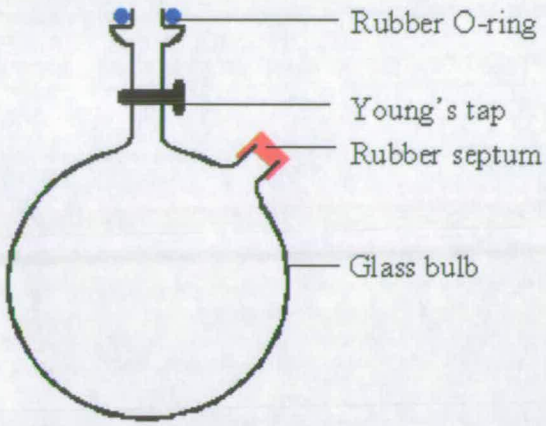


Figure 3.3 *Glass dilution bulb with side sampling port.*

3.1.4 Calculation of dilution concentrations

Dilution of standards and samples was performed by injecting known volumes of gases into dilution bulbs containing helium at known pressures (usually 10psi). Serial dilution of samples could be performed to dilute samples across a wide range. Before each gas transfer the syringe was washed with one volume of the gas being transferred. Gas concentrations were calculated using the ideal gas equation (Equation 3.1)[3].

$$pV = nRT$$ **Equation 3.1**

Where p = pressure in atmospheres
V = volume of gas (dm^3)
n = moles of gas (moles)
R = gas constant ($0.0821 dm^3.atm.K^{-1}.Mol^{-1}$)
T = temperature (K)

3.1.4.1 Typical serial dilution of gas

Each clean bulb is filled with helium (diluent gas) to a pressure of 1.6085atms (10psi above atmospheric pressure). The number of moles of helium in each bulb is calculated.

$$n = \frac{pV}{RT} \therefore n = \frac{1.6085 * 1.084}{0.082 * 293} = 0.07257 moles$$

$\therefore$ Bulb 1 = 0.07257 moles

Bulb 2 = 0.07230 moles
Bulb 3 = 0.07184 moles
Bulb 4 = 0.07217 moles

5mls gas at atmospheric pressure spiked into Bulb 1. The number of moles of sample gas at atmospheric pressure in 5ml syringe is calculated.

$$n = \frac{1*0.005}{0.0821*293} = 2.0811*10^{-4} moles$$

Therefore % mix in bulb 1 spiked with 5mls of sample gas.

$$\%\ \text{mix} = \frac{2.081*10^{-4}}{0.07257 + 2.0811*10^{-4}} * 100 = \underline{0.286\%}$$

Total gas pressure in bulb 1 after spiking with 5mls sample gas.

$$p_{spiked}V_{bulb1} = p_{initial}V_{bulb1} + p_{syringe}V_{syringe} = (1.609*1.084) + (1*0.005) \therefore p_{spiked} = 1.613 atms$$

Gas pressure in bulb 1 after 5mls removed to wash syringe.

$$p_{initial}V_{bulb1+syringe} = p_{final}V_{bulb1} \therefore p_{final} = \frac{1.613*1.084}{1.089} = 1.606 atms$$

Gas pressure in syringe used to spike 5mls of bulb1 into bulb 2.

$$p_{initial}V_{bulb1+syringe} = p_{final}V_{bulb1} \therefore p_{final} = \frac{1.605*1.084}{1.089} = 1.598 atms$$

Moles of gas in syringe used to spike 5mls bulb 1 into bulb 2.

$$pV = nRT \therefore n = \frac{pV}{RT} = \frac{(1.598*0.005)}{(0.0821*293)} = 3.326*10^{-4} moles$$

Moles of diluted sample in syringe

$$n = \left(\frac{3.324*10^{-4}}{100}\right) * 0.286 = 9.511*10^{-7} moles$$

% mix of sample in dilution bulb 2

$$\% = \left(\frac{9.511*10^{-7}}{0.0723 + 3.326*10^{-4}}\right) * 100 = 1.309*10^{-3}\%$$

These steps were used to calculate the subsequent concentrations for bulbs 3 and 4. Varying the spiking volumes (1-20ml) allowed a range of concentrations to be produced. All dilution steps and subsequent percentage mixes are detailed in Appendix C. The sequential use of the four dilution bulbs can produce dilutions in a

range of 10 orders of magnitude. In experimental terms this allows the preparation of samples containing femtogram amounts of diluted samples.

The range of different amounts of sample loaded onto the column can be further extended by varying the sample loop size (10µl - 250µl). The various dilutions and sample loops enable the loading of different amounts of sample over the extensive concentration range.

3.2 Sample loading onto the column

The different GC injection techniques[4] are all prone to errors when trying to load exact volumes of gas[5], such as operator inconsistencies. To limit errors in the loading of samples onto the GC column an injection system was employed (Figure 3.4). The injector system allowed accurate volumes (10µl-1000µl) of gas to be loaded onto the column. This consisted of an injection body and a six way gas valve with sample loop. The system was designed and built in house.

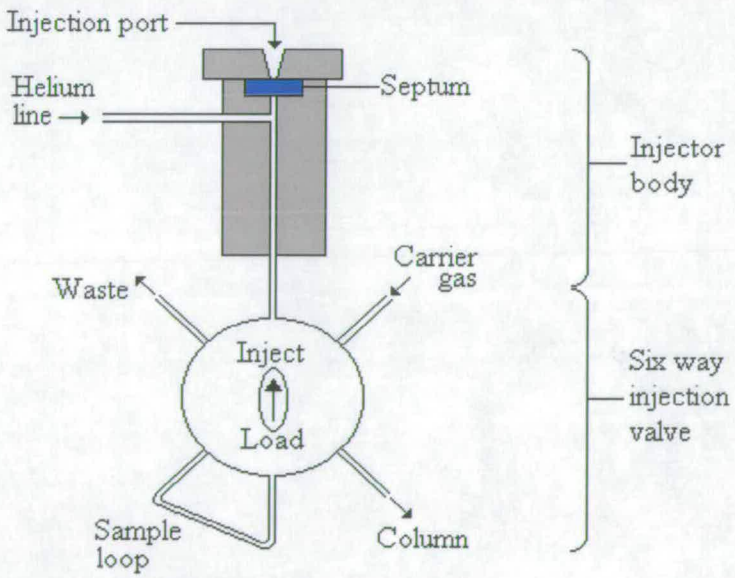


Figure 3.4 *Schematic of injection system.*

3.2.1 Injector system

The injector body consisted of an injection port which allowed gas samples to be introduction through an air tight septum into a six way valve via 1/32th stainless steel

tubing. A side connection of 1/16" stainless steel tubing was connected to the helium carrier gas line. This allowed the injection system to be flushed with helium between injections to prevent carryover of sample. The 1/16" tubing could also be connected to gas cylinders to allow direct introduction of gas samples for analysis.

The six way injection valve was fitted with loops of stainless steel tubing of different calibrated volumes (10µm-250µml). When the valve is in the load position the gas injected into the injector body flushes into the sample loop. Excess sample gas is vented out of the sample loop through the waste pipe. In the load position the carrier gas enters the valve and passes directly onto the column as shown in Figure 3.5.

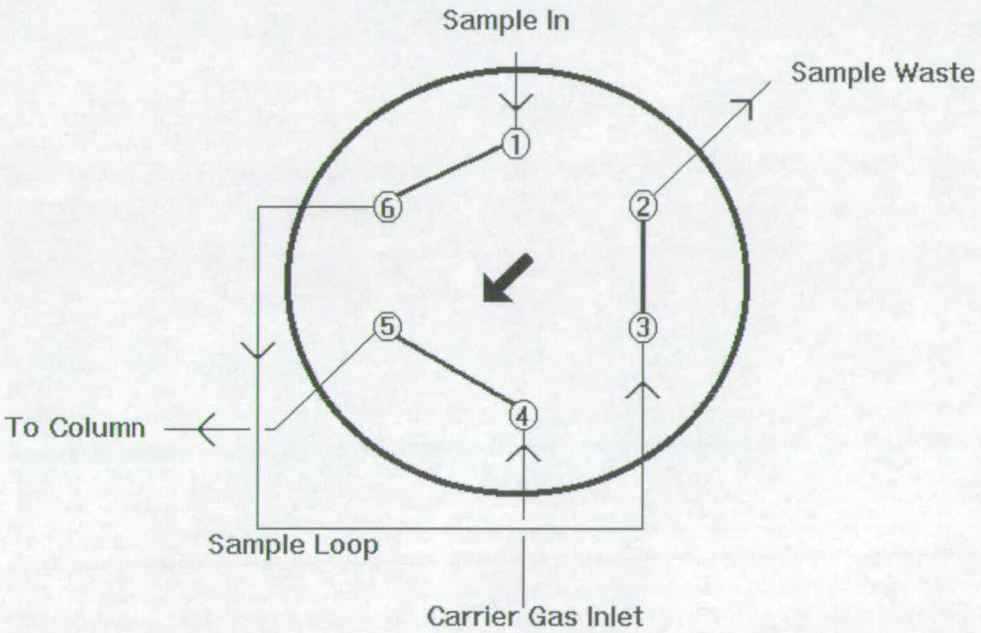


Figure 3.5: *Schematic of gas flows through the injection valve in load position.*

The sample contained within the loop is washed onto the column by switching the valve to the inject position. In the inject position the helium carrier gas enters the valve and passes through the loop before passing onto the column (Figure 3.6). This switching washes the sample gas onto the column in a tight band limiting injection effects on peak broadening. The valve is connected to the column via 1/32" stainless steel tubing using a Valco zero dead volume union (Figure 3.7).

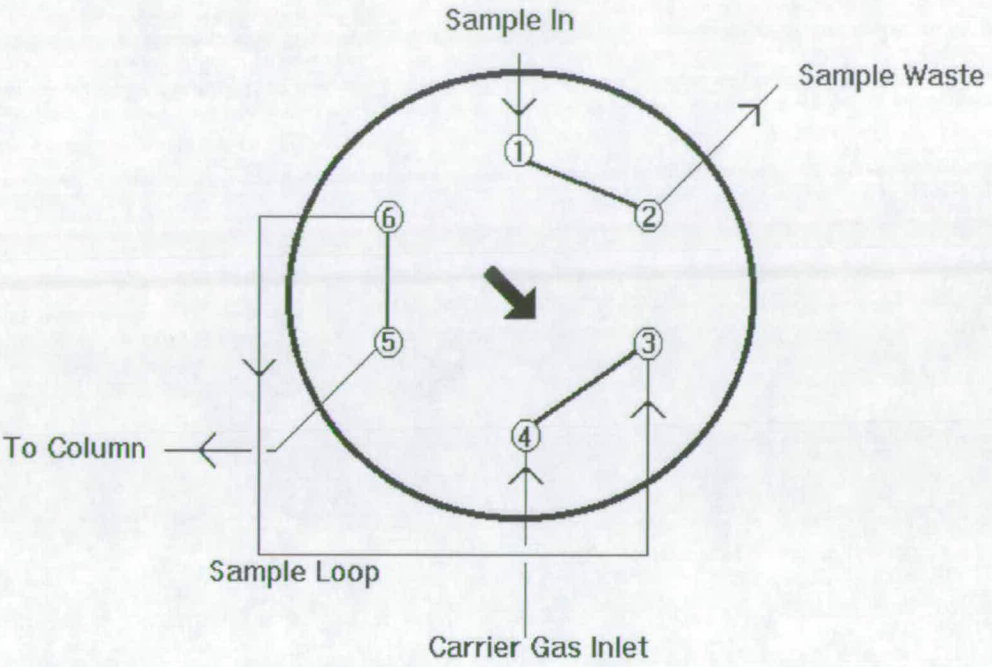


Figure 3.6: *Schematic of gas flows through the injection valve in inject position.*

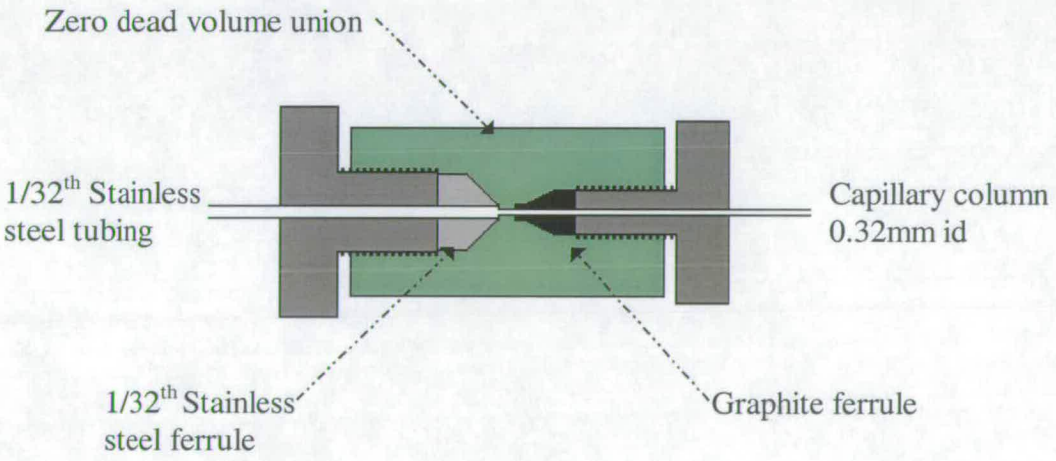


Figure 3.7 *Schematic of zero dead volume union.*

The injector system was sited directly on top of the GC oven next to the standard split/splitless injector (supplied with the instrument). The close proximity of the injector system to the GC split/splitless injector allowed temperature programming of the GC injector to raise the temperature of the injection system. This prevented any

condensation of the higher boiling point samples within the injector body (*e.g.* HCFC 123, 27.1°C).

3.2.2 Injector system testing

The injector system was tested to optimise two parameters of operation. The first was to find the minimum volume of gas required to fully saturate the sample loop to give reproducible sample loading. Secondly, the time required to flush the injector system with helium in order to prevent sample carryover between runs was investigated.

The GC was operated under the following standard conditions.

Helium carrier gas pressure	16psi
Injection volume	50µl
Column temperature	60°C
FID conditions	275°C
	30ml.min^{-1} hydrogen flow
	400ml.min^{-1} air flow
Signal attenuation	1

A series of repeat injections through the injector of various volumes (1-30ml) of a methane standard was analysed. The peak height was measured for each injection in order to find the injection volume which produced 100% FID response. The results for the analyses are shown in Figure 3.8. Maximum loop loading was shown to require at least 20mls of sample to be injected, equivalent to four full syringes.

Sample carryover was assessed by loading the sample loop using 20mls of methane, to fully saturate the loop. The injector system was then flushed with Helium (5psi) using the side port for various lengths of time (1-10mins). The helium flushing was switched off and the loop contents loaded onto the column. The height response produced by the remaining methane was then measured. The results for these experiments are shown in Figure 3.9. A minimum flushing time of 5 mins is

required to remove all sample carryover (*i.e.* to produce a response after flushing of less than the signal to noise ratio).

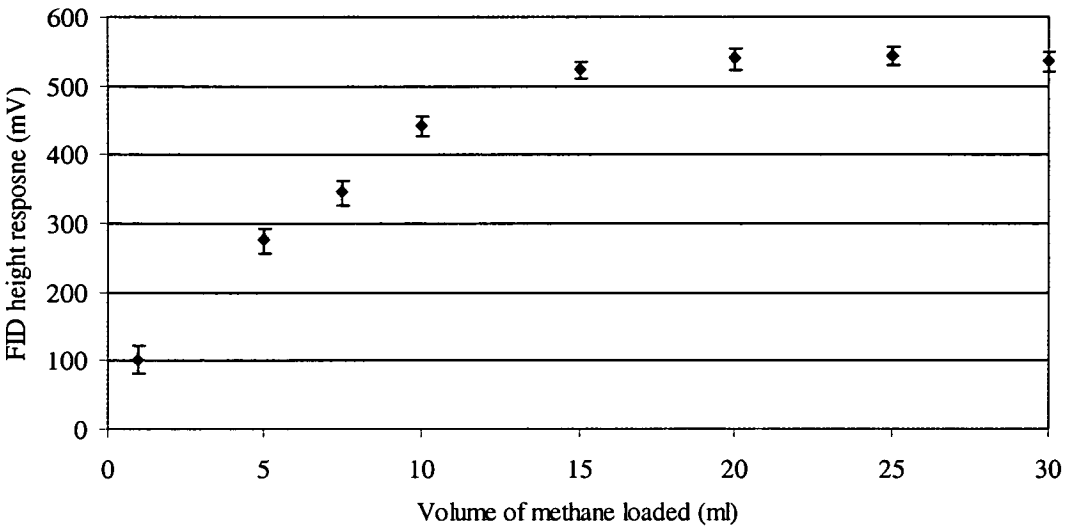


Figure 3.8 *Effect of sample volume loaded into syringe system on peak height response for 50µl methane using FID.*

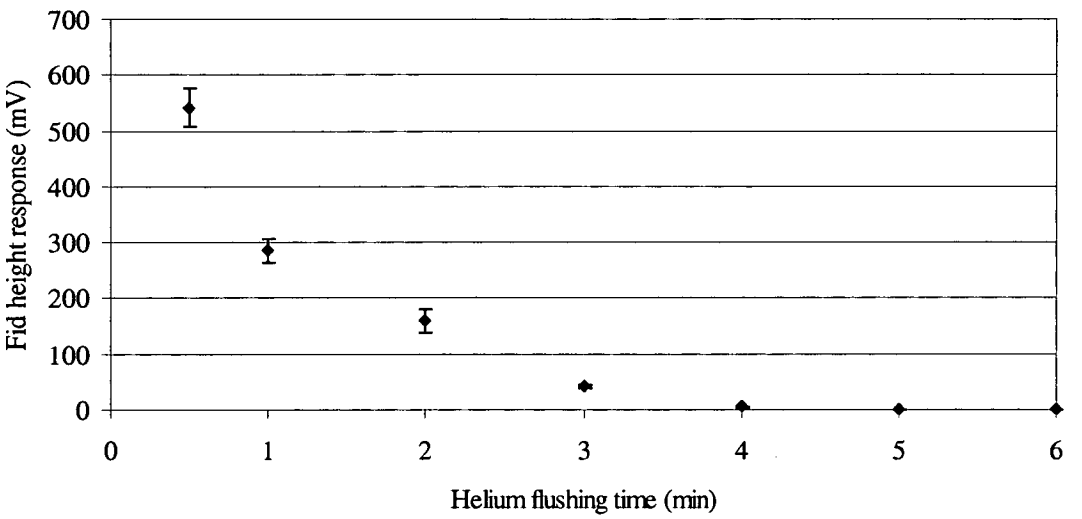


Figure 3.9 *Helium flushing time required to prevent sample carryover.*

3.3 GC separation

The GC performed in this work is separating the trace levels of synthesis by-products from each other and the major component of HFC 134a. The most important consideration in any separation protocol is the choice of column[6]. In most cases this

can be limited to investigating three or four different types evaluating the pros and cons of each in respect to the results required. Once a column type has been chosen it is necessary to optimise the conditions of operation of the column in order to attain the best possible separation. The following sections discuss the reasons for the choice of column used in this work and the steps taken to maximise its efficiency.

3.3.1 Choice of column

There have been many studies of different columns' effectivenesses in performing separations of low molecular weight halocarbons[7]. From the literature it is inferred that all the different column types currently available have certain limitations.

The family of WCOT CP-Sil CB columns have shown limited resolving power for the most volatile halocarbons[6]. As the HFC 134a samples analysed in this work are likely to contain these volatile halocarbons this group of columns do not provide the quality of separation required for this work.

Pora PLOT columns have shown good resolution for halocarbons but had excessive column bleed at higher temperatures[8]. Column bleed is the stripping of the column lining which will then elute into the detector. In the case of mass spectrometric detection column bleed swamps any signal produced by eluting components. As the experimental conditions used in these analyses involve MS this group of columns were not evaluated.

PLOT columns with aluminium oxide coated onto the fused silica (PLOT alumina) have shown excellent separation for low molecular weight halocarbons[9]. Further improvement in separation performance can be achieved by the deactivation of the alumina surface using either potassium chloride or sodium sulfate[10]. The levels of column bleed at elevated temperatures (>150°C) are not significant enough to effect mass spectrometric measurements. These columns do however show reactivity to

certain halocarbons and have been shown to induce catalytic dehydrohalogenation of some halocarbons[11,6].

The columns used in this work were PLOT alumina sodium sulfate deactivated. These columns were chosen on the basis of having excellent separation efficiencies for halocarbons and their compatibility with MS. Any effects of dehydrohalogenation on the data were closely monitored and specifically investigated using chemometrics.

3.3.2 Carrier gas choice and optimisation

The carrier gas used in this work was Helium. Helium was used as it provides high separation efficiencies at increased gas velocities when compared with nitrogen without the flammability hazards inherrent with hydrogen (see section 1.5.1.4)[12]. A series of GC analyses with detection by FID of a standard hydrocarbon mix (C_1-C_2) were performed under the following instrument conditions.

Helium carrier gas	Varied pressures
Injection volume	50µl
Column temperature	60°C
FID conditions	275°C
	30mls.min^{-1} hydrogen flow
	400mls.min^{-1} air flow
Signal attenuation	1

The carrier gas pressure was altered (5-30psi) in each analysis, all other conditions remained constant. Linear gas velocities, retention times and peak $W_{1/2h}$ were noted for all peaks in each chromatogram and are displayed in Table 3.1. These values were used to calculate the height per theoretical plate number (HETP, see Section 1.5.1.2) for each peak at each carrier gas pressure (Table 3.2).

Gas pressure	Peaks	Methane	Ethane	Ethylene	Propane	Propylene	n-Butane	Acetylene	Propyne
5 psi	Retention time (min)	9.36	10.50	11.66	15.34	18.43	24.13	32.55	38.02
	Peak height (mV)	0.67	1.44	1.31	1.96	0.15	0.97	0.64	1.71
	Peak width$_{1/2h}$ (min)	0.21	0.17	0.18	0.20	0.20	0.21	0.43	0.41
10 psi	Retention time (min)	4.48	5.60	6.22	8.19	9.86	12.89	15.09	17.32
	Peak height (mV)	1.22	3.15	2.96	4.43	2.91	3.20	1.47	2.38
	Peak width$_{1/2h}$ (min)	0.10	0.07	0.08	0.09	0.09	0.12	0.16	0.17
14 psi	Retention time (min)	3.55	3.97	4.42	5.82	7.01	9.16	12.30	14.48
	Peak height (mV)	1.92	4.94	4.74	6.83	5.66	5.03	2.34	6.21
	Peak width$_{1/2h}$ (min)	0.06	0.05	0.05	0.06	0.05	0.07	0.10	0.12
16 psi	Retention time (min)	3.22	3.60	4.00	5.27	6.34	8.28	11.12	13.10
	Peak height (mV)	2.02	5.51	4.81	6.83	9.23	4.96	2.33	6.08
	Peak width$_{1/2h}$ (min)	0.05	0.04	0.04	0.04	0.04	0.06	0.09	0.10
19 psi	Retention time (min)	2.69	3.01	3.35	4.40	5.30	6.92	9.27	10.94
	Peak height (mV)	3.08	7.95	7.71	10.53	3.12	7.67	3.75	8.72
	Peak width$_{1/2h}$ (min)	0.04	0.03	0.03	0.04	0.04	0.05	0.07	0.08
22 psi	Retention time (min)	2.31	2.58	2.87	3.79	4.56	5.97	8.01	9.45
	Peak height (mV)	3.75	9.68	9.40	12.37	5.33	8.95	4.35	11.79
	Peak width$_{1/2h}$ (min)	0.03	0.02	0.02	0.03	0.03	0.04	0.06	0.07
25 psi	Retention time (min)	2.07	2.31	2.57	3.39	4.09	5.35	7.22	8.11
	Peak height (mV)	4.67	11.63	11.31	16.63	0.88	11.70	4.99	12.17
	Peak width$_{1/2h}$ (min)	0.03	0.02	0.02	0.03	0.03	0.04	0.05	0.06
30 psi	Retention time (min)	1.73	1.94	2.15	2.84	3.42	4.45	5.92	7.05
	Peak height (mV)	6.25	14.80	14.40	18.11	3.22	12.56	6.45	13.27
	Peak width$_{1/2h}$ (min)	0.02	0.02	0.02	0.02	0.03	0.03	0.04	0.06

Table 3.1 *Peak data for Van Deemter plots using standard hydrocarbon mix with a range of helium carrier gas pressures.*

	Ethane	Ethylene	Propane	Propylene	n-Butane	Acetylene	Propyne
5 psi	210.04	56.74	10.43	4.58	1.88	3.03	1.85
10 psi	37.08	17.11	5.19	2.53	1.72	2.02	1.56
14 psi	103.87	28.16	5.20	2.01	1.55	1.13	0.99
16 psi	69.94	18.25	3.55	1.44	1.10	0.89	0.79
19 psi	73.21	20.61	4.51	1.71	1.21	0.97	0.92
22 psi	66.07	15.88	3.82	1.67	1.25	0.99	0.93
25 psi	64.00	15.95	3.16	1.73	1.27	0.99	1.02
30 psi	60.03	14.61	3.88	1.98	1.29	1.00	1.09

Table 3.2 *HETP(mm) for standard hydrocarbon mix at various helium carrier gas pressures pressures.*

Van Deemter plots were drawn for propyne, acetylene, n-butane and propylene. Error bars of 5% were assigned to the values calculated for HETP because of measurement errors and between sample variations (see Section 4.1). Van Deemter plots were not drawn for peaks eluting before propylene because these peaks do not give a good representation of the chromatographic performance of the column as they elute too quickly. From the Van Deemter plots for propyne (Figure 3.10) and propylene, n-butane and acetylene (Figure 3.11) it can be seen that column performance is optimised at a linear velocity of approximately 26cm.s^{-1} (gas pressure 16psi).

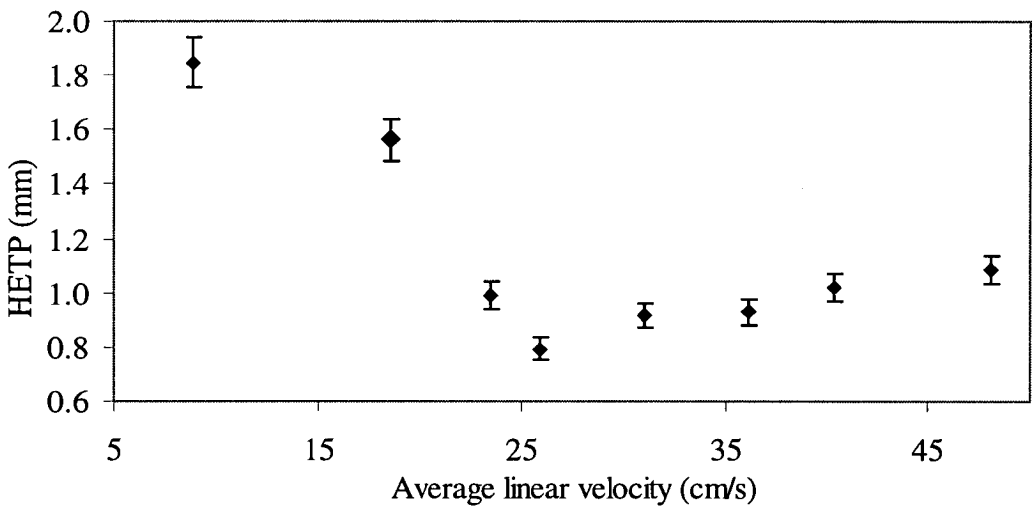


Figure 3.10 *Van Deemter plot for propyne*

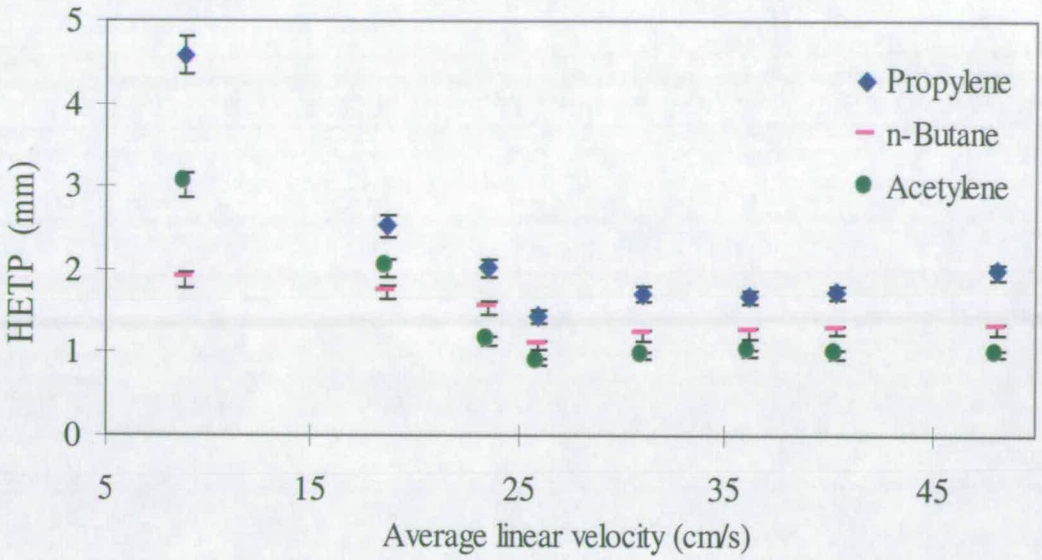


Figure 3.11 *Van Deemter plots for propylene, n-butane and acetylene*

3.4 Detection of components

There are a number of different detectors available for use with GC equipment which have been used to study the compounds of interest in this work[8,13]. Due to equipment constraints the detectors evaluated and used in this work were the Flame Ionisation Detector (FID) and the Electron Capture Detector (ECD).

3.4.1 FID

The operating parameters of the FID are gas flow rates for hydrogen and air, and temperature of the detector body. The gases are used to produce the flame which eluting samples are combusted in to form ions for detection[14]. The effect of altering these gases has not been studied in this work. The detector is held at a high temperature (10^{o} > than the maximum column temperature) to prevent condensation within the detector body of eluting components and compounds formed in the flame.

A series of analyses were performed on a sample of HFC 134a to determine the effect of detector temperature on signal output. The analyses were performed under the following instrument conditions.

Helium carrier gas pressure 16psi

Injection volume	50µl
Column temperature	120°C
FID conditions	Varied detector temperature
	30mls.min^{-1} hydrogen flow
	400mls.min^{-1} air flow
Signal range	5

All chromatographic conditions were kept constant as the temperature of the detector was altered from 150°C to 380°C. Each temperature experiment was repeated in triplicate. The results were measured as the peak height response and are displayed in Figure 3.12. The graph shows that increases in the temperature above 200°C do not have any significant effect on the signal output. From these results it can be concluded that increases in temperature over 200°C do not significantly alter the formation of ions. As there are no apparent gains from altering the detector temperature it will be operated at the manufacturer's recommended temperature of 275°C[15].

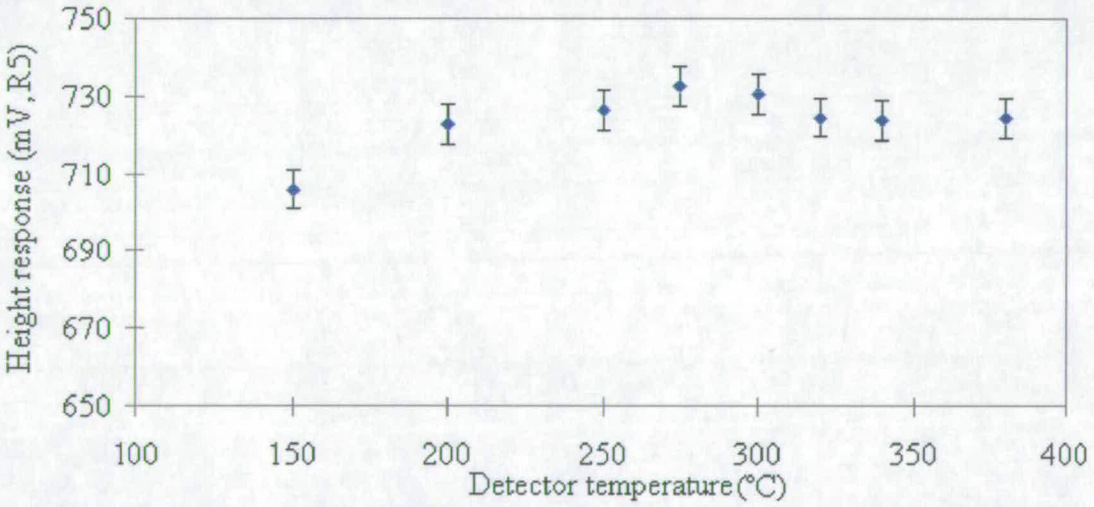


Figure 3.12 *Height response of HFC 134a with detection by FID at various detector temperatures*

3.4.2 ECD

ECD response is affected both by detector temperature and make up gas flow. Increasing the make up gas flow increases the linear range for more concentrated samples but lowers the sensitivity. Increasing the temperature increases the

sensitivity of the detector but leads, in time, to the oxidation of the collector electrode and subsequent reduction in detector performance[16].

Make up gas and detector temperature are the two conditions which can be altered when using the ECD. Manufacturers specifications state the range of values for make up gas are 30-60 ml min^{-1} with an operating temperature of 300-350°C[16]. The operating conditions will be set to cover these ranges and extend both above and below.

3.4.2.1 Make up gas

A range of gas flow rates from 20ml min^{-1} to 100ml min^{-1} were investigated to find the optimum flow rate in terms of sensitivity and signal stability. A series of repeat analyses were performed on a sample of CFC 115 (chloropentafluoroethane) under the following conditions. The sample of CFC115 was diluted once to a concentration of 0.2% in helium (µg analysed)

Helium carrier gas pressure	16psi
Injection volume	50µl
Column temperature	100°C isothermal
ECD conditions	350°C
	Varied nitrogen flow
Signal range	5

The make up gas flow rate was altered (5-60ml.min^{-1}) between analyses with each flow rate being repeated five times. Figure 3.13 shows the relationship between gas flow rate and detector response. %RSD values for the five repeat analyses were less than 2.6% at all gas flow rates. From the graph it can be seen that increases in the nitrogen make-up gas flow rate decreases the ECD response. This is because increased flow rates washes the sample through the detector quicker than at lower flow rates. As the sample spends less time in the detector assembly it has less time to

interact with nitrogen ions formed in the detector causes a decrease in signal response.

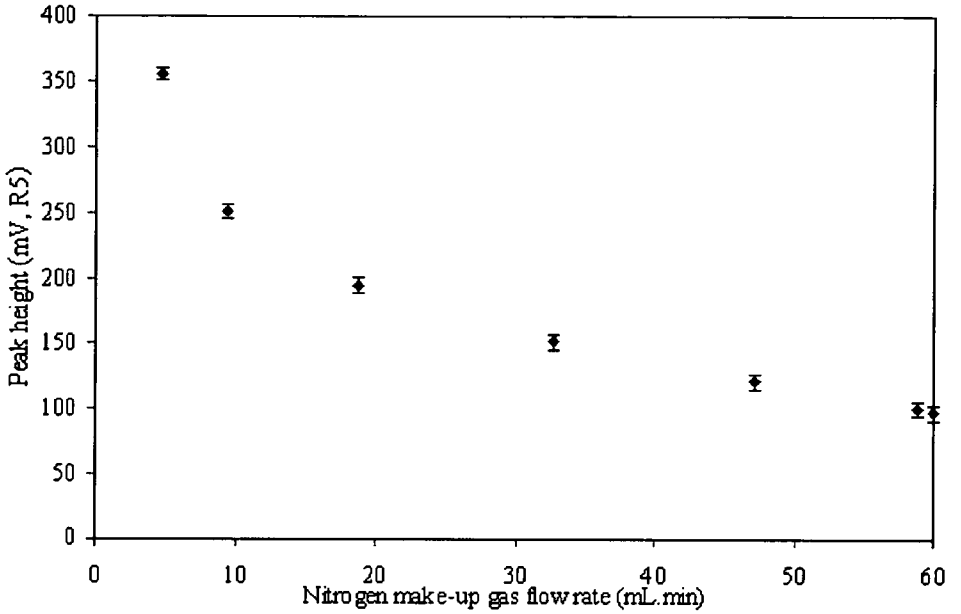


Figure 3.13 *Effect of varying nitrogen make-up gas flow rate on ECD signal response to CFC 115*

3.4.2.2 Detector temperature relationship

The operating temperature of the ECD was evaluated over the range 200°C-375°C. Temperatures in excess of 350°C are not recommended by manufacturers as these increase the rate at which the electrodes are oxidised reducing the lifetime of the detector. A series of repeat analyses of HFC 134a were performed using the following conditions.

Helium carrier gas pressure	16psi
Injection volume	50µl
Column temperature	50°C for 30mins, increase to 100°C at 20°.min^{-1} hold for 20mins, increase to 180°C at 10°.min^{-1} hold for 10 mins
ECD conditions	Varied detector temperature 30mls.min^{-1} nitrogen flow
Signal range	1

Using the column temperature program allowed the separation of the by-products contained within the HFC 134a sample analysed. This allowed the effect of the detector temperature to be monitored for a range by-products present in the sample of HFC 134a. The detector signal was allowed to stabilise between increases and decreases in the temperature until a steady state signal without baseline drift was observed. The peak heights for each component were measured. The values obtained for the thirteen components gave a good representation of the effect of ECD temperature on the relative sensitivity of the detector. Each temperature value was repeated in duplicate.

Data representing the temperature dependant ECD response are displayed in Table 3.3 and illustrated in Figures 3.14-3.20. Each data point represented in the plots corresponds to the peak height value measured from the chromatograph. The data are plotted as the natural logarithm of the peak height as a function of the inverse temperature *i.e.* Arrhenius fashion[17].

	Detector temperature in °C											
Peak No.	**200**	**200**	**250**	**250**	**300**	**300**	**325**	**325**	**350**	**350**	**375**	**375**
1	0.29	0.35	0.73	0.70	1.61	1.45	2.44	2.83	3.33	3.39	3.03	3.06
2	1.00	0.98	1.48	1.49	2.17	2.19	2.55	2.56	2.85	2.81	3.25	3.26
3	0.25	0.20	0.31	0.19	0.21	0.20	0.22	0.25	0.29	0.30	0.52	0.48
4	3.03	2.91	4.11	4.17	5.65	5.81	6.34	6.31	7.01	6.43	7.87	8.04
5	0.00	0.00	0.00	0.00	0.00	0.076	0.104	0.095	0.113	0.103	0.12	0.16
6	0.29	0.37	0.39	0.24	0.51	0.564	0.82	0.80	1.08	1.10	1.48	1.56
7	30.5	31.9	43.3	42.97	59.2	60.8	72.9	68.21	85.2	97.3	101.2	88.2
8	2.19	2.42	2.95	2.88	4.47	4.68	5.60	5.19	7.27	7.43	6.98	4.89
9	549.5	576.0	681.2	636.1	793.1	826.7	944.1	833.9	951.1	966.8	1069	1048
10	0.11	0.12	0.41	0.51	2.22	2.19	4.26	4.58	5.76	6.49	12.2	11.8
11	6.28	6.77	11.4	11.69	24.3	22.6	30.3	27.2	34.4	37.3	43.3	44.3
12	15.2	13.3	16.3	16.4	24.5	27.7	36.9	26.5	26.8	31.37	24.4	24.7
13	0.00	0.00	0.00	0.00	0.24	0.28	0.48	0.31	0.61	0.83	0.58	0.69

Table 3.3 *HFC 134a components peak heights analysed over a range of ECD temperatures.*

The slopes of the lines are proportional to the activation energy for ion fragmentation or the enthalpy of molecular anion formation. Negative slopes indicate a dissociation

ion fragmentation electron attachment mechanism, whereas a positive slope indicates electron attachment by a resonance mechanism[18].

Peaks 3 and 6 in Figure 3.14 show a negative to positive slope transition as the temperature is lowered (>300 and >250°C respectively), indicating a transition from dissociative to attachment ionisation mechanism when the temperature is lowered. Increasing the temperature range would allow this hypothesis to be further tested.

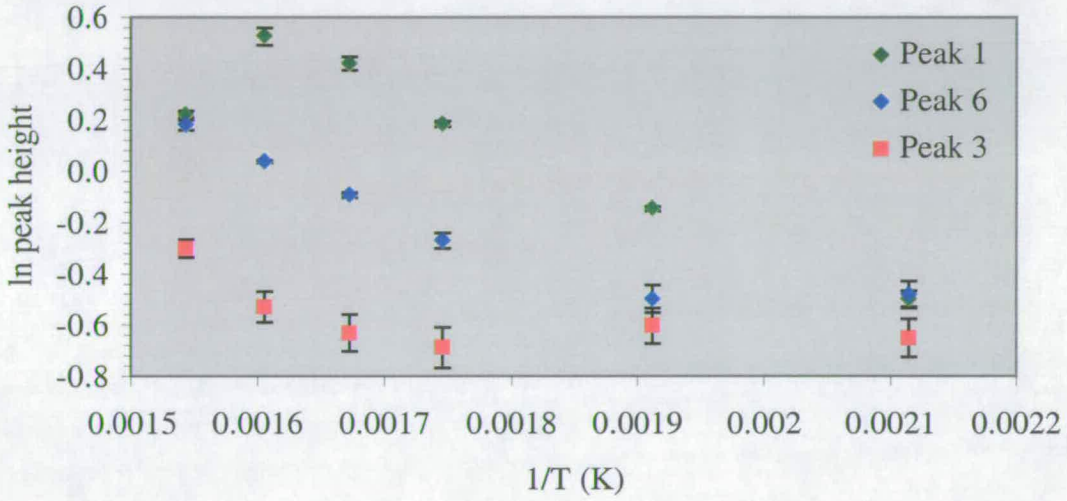


Figure 3.14 *ln (detector response) as a function of inverse temperature from 200 to 375°C for peaks 1, 3 and 6*

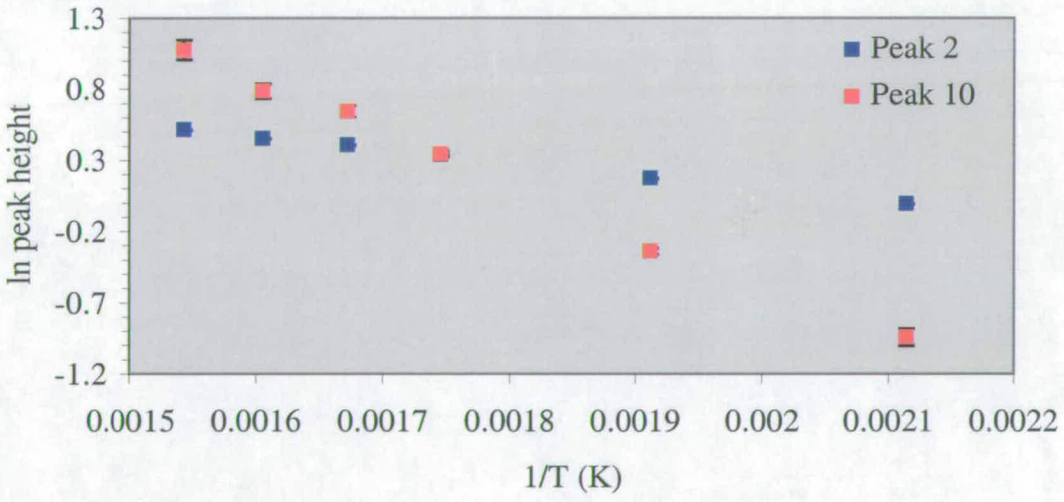


Figure 3.15 *ln (detector response) as a function of inverse temperature from 200 to 375°C for peaks 2 and 10*

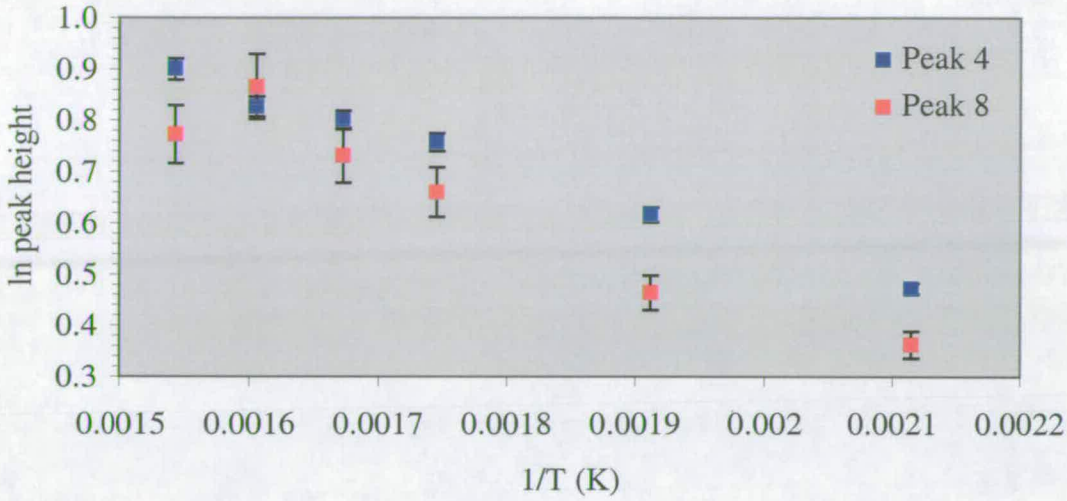


Figure 3.16 *ln (detector response) as a function of inverse temperature from 200 to 375°C for peaks 4 and 8*

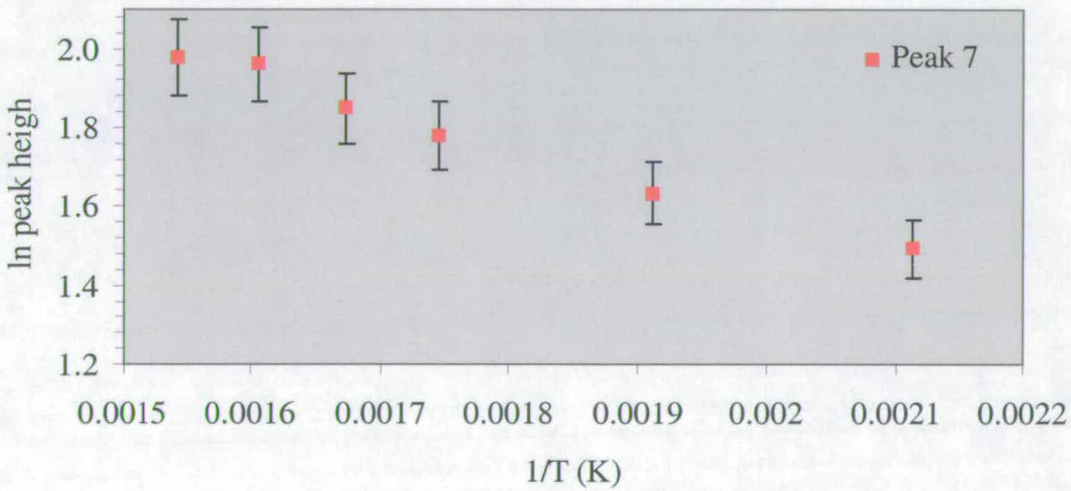


Figure 3.17 *ln (detector response) as a function of inverse temperature from 200 to 375°C for peak 7*

Figures 3.15, 3.16, 3.17, 3.19 and 3.20 show that peaks 2, 4, 7, 9, 10 and 11 all have negative slopes, indicating that electron attachment occurs by the dissociative mechanism. As seen in Figures 3.12, 3.14 and 3.18, peaks 1, 8 and 12 have complex Arrhenius plots. The plots have a positive slope at high temperature which changes to a negative slope at lower temperatures, indicating a mixed mechanism of ionisation. This ionisation profile would involve electron attachment at higher

temperatures and dissociative ionisation at lower temperatures. Further analyses at higher temperatures (>375°C) are required to fully test this hypothesis.

The ionisation processes for peaks 5 and 13 (Figure 3.18) are difficult to assign as no values were obtained for temperatures below 325°C and 300°C respectively. However, as the height response has dropped below LOD (*i.e.* values off the bottom of the graph) this would indicate a continuing dissociative ionisation mechanism.

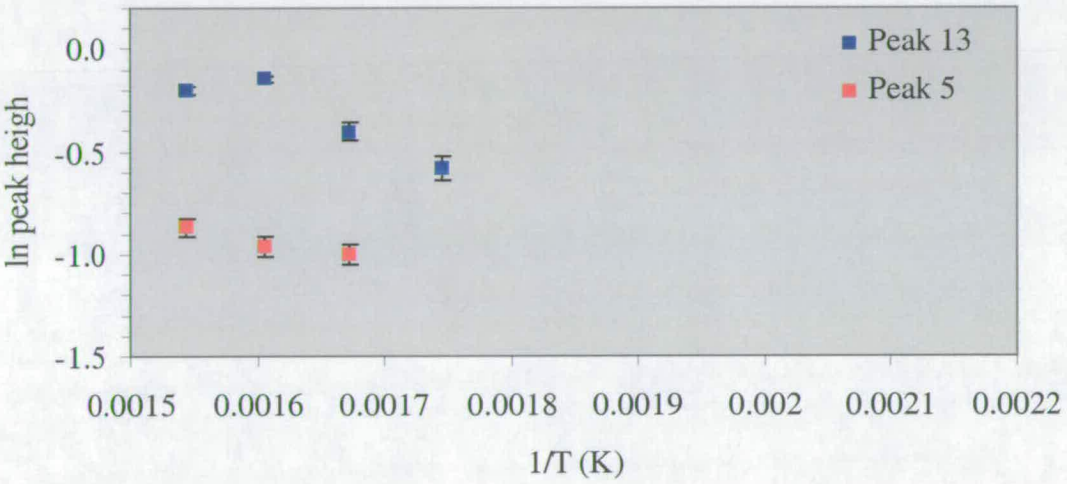


Figure 3.18 : *ln (detector response) as a function of inverse temperature from 200 to 375°C for peaks 5 and 13*

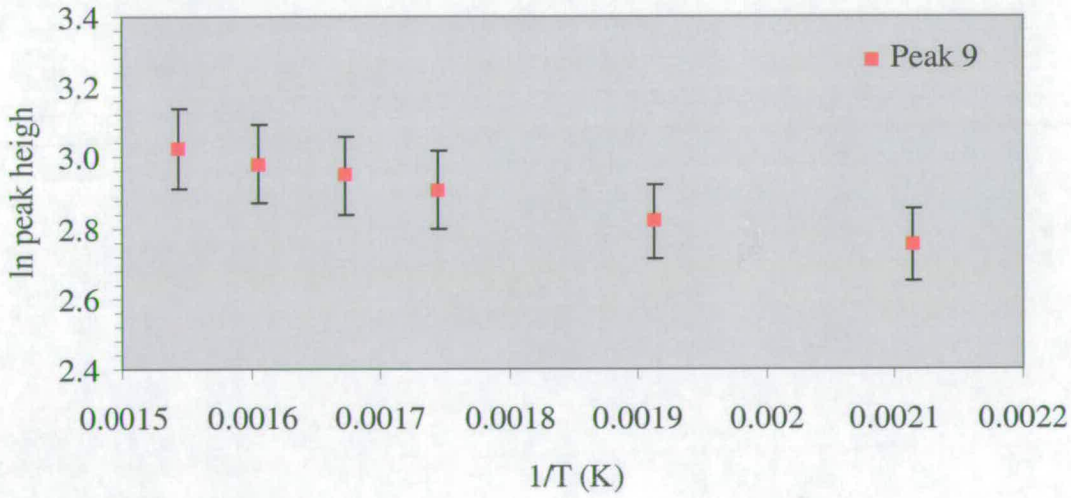


Figure 3.19 *ln (detector response) as a function of inverse temperature from 200 to 375°C for peak 9*

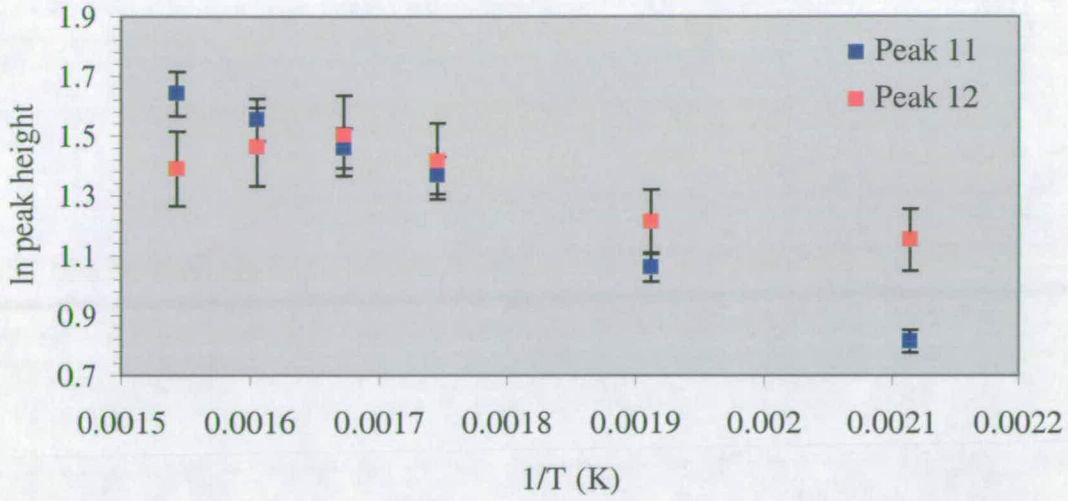


Figure 3.20 *ln (detector response) as a function of inverse temperature from 200 to 375°C for peaks 11 and 12.*

The results from these sets of analyses show higher detector temperatures lead to better sensitivity for the majority of components. However, because of oxidation of the detector electrode at temperatures > 350°C it was decided to operate the ECD at 350°C to give the best sensitivity while avoiding excess wear on the ECD.

3.5 Mass Spectrometry

The GC-MS instruments used in this project were a Finnigan 4600 GC-MS (sited at Edinburgh University) and a Kratos Concept 15 (sited at ICI C&P, Runcorn).

3.5.1 Finnigan 4600 GC-MS

The Finnigan 4600 GC-MS consisted of a Perkin Elmer GC fitted with a sample injector system (as described in section 3.2.1). The PLOT alumina capillary column was fed from the GC through a transfer oven into the ionisation source. The transfer oven and ionisation source were held at a temperature of 150°C to prevent condensation of the column eluant. The capillary column terminated a couple of mm from the electron beam. The carrier pressure used corresponded to a flow rate of less than a couple of $ml.min^{-1}$ negating the need for any special interface between the gas

chromatograph and the mass spectrometer[19]. The ionisation source is pumped to a pressure of 10^{-6} Torr.

Ionisation of column eluant was performed using electron impact ionisation. The electron beam voltage was set at 70eV to give optimised ionisation efficiencies[20]. The analyser assembly was set to scan over the mass range 30-350Da with a scan time of one second. The analyser was calibrated on a daily basis using the standard mass spectrometer calibrant, perfluorokerosene. The electron multiplier was set at 1.5kV to give a level of sensitivity with an acceptable level of background noise. The analyser assembly and electron multiplier were pumped to a pressure of 10^{-7} Torr.

The GC was programmed directly using the instrument keypad. The mass spectrometer was controlled using a workstation computer running the INCOS data handling software. Data was viewed using the INCOS data system fitted with a printer. Data files were stored on magnetic tape.

3.5.2 Kratos Concept 1S GC-MS

The Kratos Concept 1S consisted of a Hewlett Packard 5890 GC connected to a Kratos Concept 1S mass spectrometer. Sample injection was through the split injector with a split ratio of 20:1 (*i.e.* $1/20^{th}$ of the injected volume passes onto the column). The PLOT alumina column is fed from the GC into a glass transfer line. The transfer line terminates a couple of mm from the electron beam in the ionisation source. The transfer line and ionisation source are held at a temperature of 150°C. The ionisation source was pumped to a pressure of 10^{-6} Torr.

Ionisation of column eluant was achieved using electron impact ionisation at voltage of 70eV. The instrument is a double focusing reversed geometry sector mass spectrometer. The analyser was set to scan from 30-350Da at one scan per second. The analyser was tuned and calibrated automatically using perfluorokeroene once a

day. The electron multiplier was operated at 1.5kV. The analyser region and electron multiplier were operated under vacuum (10^{-7} Torr).

The GC was programmed and operated using the instrument keypad. The mass spectrometer was controlled using a PC running the MASPEC II mass spectrometer software (produced in-house by ICI). Data handling and manipulation was performed using the MASPEC II software. Data storage was on the hard disk of the PC and hard copies were produced using a laser printer.

3.6 Data acquisition and handling

Data acquisition for the GC equipment was performed using a PC running the 1020 Perkin Elmer chromatographic data handling program. The Finnigan and Kratos mass spectrometers acquired data using the INCOS and MASPEC II data systems respectively. The file formats produced using these different data acquisition systems were not compatible. The direct transfer of data files, *i.e.* the values for each point in the different chromatograms, was not possible using the 1020 software and the INCOS data systems. The information required from each chromatographic trace whether produced by FID, ECD or MS was always the same. It consisted of a set of values which represented the amount present of each component as detected by the system used. This value could be calculated as either peak height or peak area.

A major problem in all three data handling systems is that the data format cannot be exported to the chemometrics package used. The ideal situation would involve transferring all the data points which describe the chromatographic profile for an analysis to the statistical package. Instead only the values for the peak heights/areas were manually input into the statistical package spreadsheet. This leads to alterations of the raw data through the integration process of the different data systems and adds extra variation to the data.

3.6.1 Baseline noise

Baseline noise levels were measured for each sample analysed by looking at a number of chromatographic traces for both ECD and FID. The noise levels were measured before and after elution of the HFC 134a peak. The time point at which each measurement was taken varied between traces to avoid peaks and baseline drift due to temperature gradients. Table 3.4 displays the results of these measurements (all values are subject to a measurement error of ± 0.002mV). From the Table the values for the noise levels before and after the elution of HFC 134a can be approximated to (0.051 ± 0.02)mV. The value for peak height recognition (3:1 signal:noise ratio) is set at (0.150 ± 0.02mv).

Table 3.4 : *Measurement of background noise levels*

Sample	File number	Pre HFC 134a noise (mV)	Post HFC 134a noise (mV)
UK	01048R05	(95.706-95.655) = **0.031**	(96.651-96.600) = **0.031**
USA	02048R02	(95.910-95.858) = **0.032**	(97.070-97.120) = **0.030**
Japan	03048R03	(96.151-96.101) = **0.030**	(97.679-97.627) = **0.032**
Ashai	07048R03	(95.149-95.098) = **0.031**	(95.898-95.850) = **0.028**
Atochem	07048R02	(96.178-96.126) = **0.032**	(97.740-97.690) = **0.030**
Dupont	08048R02	(96.131-96.082) = **0.029**	(96.743-96.694) = **0.029**
Showa	08048R01	(97.519-97.466) = **0.033**	(96.051-96.000) = **0.031**
Hoechst	09048R05	(95.209-95.159) = **0.030**	(96.431-96.380) = **0.031**
		Average value **0.031**	Average value **0.030**

3.6.2 Baseline drift

The baseline drift was measured for both the FID and ECD. The multi-step temperature program used to separate the samples was run without the injection of a sample. The chromatographic trace was acquired over the run time and the level of baseline drift was monitored. The chromatographs were considered in five regions. Region 1 was the isocratic initial temperature of 50°C. Region 2 was the first temperature step from 50-100°C at $20^{\circ}.min^{-1}$. Region 3 was the second isocratic temperature region 100°C for 20mins. Region 4 was the second temperature step from 100-180°C at $10^{\circ}.min^{-1}$. Region 5 was the final isocratic temperature region 180°C for 5mins.

The results for these analyses are displayed in Table 3.5. Region 1 shows no baseline drift for both FID and ECD analysis. The regions of temperature increase show the greatest amount of baseline drift, with the amount of drift being proportional to the rate and amount of temperature increase. Regions directly after temperature increase show low levels of stability.

Table 3.5 *Levels of baseline drift over the course of the multi-step temperature program for FID and ECD analysis*

Detection	Region 1 (50°C)	Region 2 (20°.min^{-1})	Region 3 (100°C)	Region 4 (10°.min^{-1})	Region 5 (180°C)
FID	0mV.min^{-1}	0.4mV.min^{-1}	0mV.min^{-1}	0.2mV.min^{-1}	-0.05mV.min^{-1}
ECD	0mV.min^{-1}	0.2mV.min^{-1}	0mV.min^{-1}	0.1mV.min^{-1}	-0.02mV.min^{-1}

3.6.3 Peak measurements

As discussed earlier the data collection system for the GC with detection by FID and ECD was the 1020 software. A representative selection of peaks from the replicate analyses of the ICI UK sample of HFC 134a were used to validate the method of peak measurement. The measurements were performed in one of three ways; automatic peak integration with a moving baseline, automatic peak height measurement with a moving baseline and manual peak height measurement using the post processing view option (the software was not supplied with a manual peak integration option).

The results for these analyses are shown in Table 3.6 were the different values obtained by the different peak measurements are displayed; automatic area measurements, automatic height measurements and manual height measurements. The values quoted in the final column show the percentage relative standard deviation for the measurement techniques. These values show, on average, the most reproducible method of peak size calculation is the manual height measurement. Although the chemometric techniques used might have been able to deal with the lower quality data it was decided to use the more time consuming method of manual measurement in order to obtain the best quality data. The manual measurement of

peaks also allowed the close inspection of the data and identification of peaks 'overlooked' by the automatic integration and removal of 'false peaks'.

Table 3.6 *Peak information obtained from the different measurement techniques for FID replicate analyses of ICI UK produced HFC 134a.*

	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6			
Automatic area							Average	Std Dev.	**%RSD**
Peak 1	26448	27724	59888	20288	20832	44696	33313	15745	**47**
Peak 2	20080	19249	18676	26574	16562	25052	21032	3911	**19**
Peak 3	32268	31194	29627	27015	25530	29180	29136	2521	**9**
Peak 4	7128	6511	6484	6082	6790	7163	6693	417	**6**
Peak 5	178664	291544	223068	194346	189774	189096	211082	42157	**20**
Peak 6	52614	39177	40997	37471	37670	37514	40907	5897	**14**
Automatic height									
Peak 1	3.7188	3.7959	3.4408	3.0319	2.9953	3.0773	3.3433	0.3589	**11**
Peak 2	2.1606	2.1673	1.9874	1.6441	1.5995	1.7625	1.8869	0.2532	**13**
Peak 3	3.9767	3.8678	3.4500	2.8763	2.5826	2.9446	3.2830	0.5695	**17**
Peak 4	0.8666	0.8674	0.7626	0.6500	0.6362	0.7000	0.7471	0.1029	**14**
Peak 5	6.7773	7.9342	5.0843	7.0492	6.9501	7.0352	6.8051	0.9349	**14**
Peak 6	2.8936	1.6645	1.7822	2.5100	2.4456	2.5971	2.3155	0.4851	**21**
Manual peak height									
Peak 1	3.723	3.791	3.410	3.020	3.247	3.077	3.3780	0.3245	**10**
Peak 2	2.154	2.184	1.971	1.620	1.799	1.717	1.9075	0.2332	**12**
Peak 3	3.969	3.865	3.721	3.276	2.793	2.897	3.4202	0.5055	**15**
Peak 4	0.813	0.865	0.833	0.651	0.717	0.681	0.7600	0.0885	**12**
Peak 5	7.570	8.077	5.045	7.016	6.935	7.636	7.0465	1.0680	**15**
Peak 6	3.048	1.847	2.272	2.529	2.417	2.692	2.4675	0.4042	**16**

References

[1] de Zeeuw, J., Zwiep, D. and Marinissen, *International Laboratory*, **September**, p12J-12P, (1996)

[2] *Chromatogrphic Products for Analysis and Purification*, Supelco, United Kingdom, p84-85, (1998)

[3] Atkins,P.W., *Physical Chemistry*, Oxford University Press, England, 2nd ed., p21, (1984)

[4] Grob, K., *Anal. Chem.*. **66**, N20, p1009A-1018A, (1994)

[5] Rood, D., *A Practical Guide to the Care, Maintenance and Troubleshooting of Capillary Gas Chromatographic systems*, Huthig, Heidleberg, (1991)

[6] Braithwaite, A. and Smith, F.J., *Chromatographic Methods*, 4th ed., Chapman and Hall, London, p156, (1985)

[7] Sturrock, G.A., Simmonds, P., Nickless, G., Zweip, D., *J. Chromatogr*, **648**, 423-431, (1993)

[8] Simmonds, P.G., O'Doherty, S., Nickless, G., Sturrock, G.A., Swaby, R., Knight, P., Ricketts, J., Woffendin, G. and Smith, R., *Anal. Chem.*, **67**, 717-723 (1995)

[9] de Nijs, R.C., de Zeeuw, J., *J. Chromatogr*, **279**, 41-48, (1983)

[10] Do, L. and Raulin, F., *J. Chromatogr*, **514**, p65-69, (1990)

[11] Noy, T., Fabian, P., Borchers, R., Cramers, C., Rijks, J., *Chromatographia*, **26**, 149-156, (1998)

[12] Fowlis, I.A., *Gas Chromatography*, 2nd ed., John Wiley and Sons, Chichester, England, p12-14, (1995)

[13] Bachmann, K. and Reineke, F.J., *J. Chromatogr*, **323**, 323-329, (1985)

[14] Schomburg, G., *Gas Chromatography*, VCH Publishers, New York, p116, (1990)

[15] *Perkin Elmer Autosystems GC: Operators Manual,* Perkin Elmer, UK, Chapter 7, (1990)

[16] *Perkin Elmer Autosystems GC: Operators Manual*, Perkin Elmer, UK, Chapter 8, (1990)

[17] Atkins,P.W., *Physical Chemistry*, Oxford University Press, England, 2nd ed., p935-938, (1984)

[18] Sousa, S.R. and Bialkowski S.E., *Anal. Chem.* **69**, p3871-3878, (1997)

[19] Chapman, J.R., *Practical Organic Mass Spectrometry*, 2nd ed., John Wiley and Sons, Chichester, England, p34-37, (1993)

[20] McLafferty, F.W., *Interpretation of Mass Spectra*, University Science Books, Mill Valley, Califonia, 3rd ed., p5, (1980)

Chapter 4

Chromatographic Profiling Results

4.0 Introduction to chapter

This chapter details the results obtained for the analyses of samples of HFC 134a. The chapter is divided into sections as follows. Section 4.1 details the results of performance testing of the experimental equipment. Section 4.2 details the method development used to determine a temperature program for the separation of the by-products of HFC 134a. It also presents the results obtained using this separation method with the different detection systems coupled to the GC. Section 4.3 details the results of the chromatographic profiles obtained for the samples of HFC 134a and shows how the required information for chemometric analysis was extracted. Section 4.4 presents the results for the GC-MS identification of the by-products contained in the different samples of HFC 134a and how these results were interpreted. Section 4.5 presents the data obtained from analysing mixed samples of HFC 134a and how these can be used in the comparison of the different samples. Section 4.6 details the transfer of data into a format compatible with the multivariate statistical package used in this work.

4.1 Performance testing of experimental equipment

To ensure that the results obtained from the analysis of HFC 134a samples are an accurate representation of the by-products present in the samples it is necessary to monitor the performance of the equipment. Close monitoring of the equipment ensures that any changes in the profiles of different samples are because of differences within the samples and are not due to effects of the equipment[1].

4.1.1 GC column performance

The performance of the GC columns was monitored in two ways. The first was to inspect the chromatographs obtained for HFC 134a samples for any abnormalities. These abnormalities included increases in baseline noise or drift, large changes (>1 minute) in the retention time of the leading edge of the peak for HFC 134a or increases in peak broadening and tailing.

The second, more precise way, was to analyse a standard sample of low molecular weight hydrocarbons under a set of standard conditions (see below) and to measure the separation efficiency (Section 1.5.1.2). Once these measurements were obtained they were stored as a record (see Appendix D) and at various time points (approximately once a month) the standards were re-run and checked against the original values. Any deterioration in the column performance could therefore be monitored. The results for the performance testing of the two PLOT Alumina sodium sulphate deactivated columns used in this work are displayed in Appendix D. A selection of those results for PLOT column 1, covering the time period over which this work was performed, is shown in Table 4.1. The %RSDs show the variation of calculated resolution values with time. Both columns' separation performances did not alter by any significant level over the time course of the analyses performed in this work. That is, the repeated use of the columns did not significantly alter their resolving power.

Standard conditions:	column temperature	60°C isocratic for 20 mins
	carrier gas	Helium 16psi
	FID	275°C
		40mls.min^{-1} hydrogen
		400mls.min^{-1} air

Table 4.1 *Column performance monitoring for PLOT column 1 measured as peak resolution values for standard low molecular weight hydrocarbon mix where Peak 1 = methane, Peak 2 = ethane, Peak 3 = ethene, Peak 4 = propane, Peak 5 = propene, Peak 6 = n-butane, Peak 7 = ethyne and Peak 8 = propyne.*

Date	Peak 1-2	Peak 2-3	Peak 3-4	Peak 4-5	Peak 5-6	Peak 6-7	Peak 7-8
Jun-96	3.85	4.51	12.83	10.82	16.44	17.04	9.88
Oct-96	4.10	4.52	14.97	11.27	18.30	18.57	10.58
Jan-97	4.04	4.79	15.70	10.12	17.38	18.57	10.43
Mar-97	3.92	4.52	12.76	10.50	15.98	16.62	9.44
Oct-97	3.99	4.36	12.93	11.29	16.62	17.98	10.08
Feb-98	3.95	4.55	14.49	11.48	16.57	17.14	9.58
Jul-98	3.97	4.38	14.12	10.59	17.11	17.59	9.76
%RSD	2.0	3.1	8.3	4.6	4.5	4.3	4.3

4.1.2 Monitoring of FID performance

The signal output from the FID was monitored to avoid changes in by-product peak heights occurring due to equipment error as opposed to differences in the samples.

The monitoring was performed using an external standard of methane. 50µl of methane was analysed under a set of standard conditions when the FID was initially switched on and at intervals throughout the continuous use of the FID (approximately once a week). All the results for this performance monitoring are displayed in Appendix E. The Shewhart chart[2] in Figure 4.1 displays a selection of the test results representing monthly values over the period of this research project.

The methane standard was loaded and chromatographed following the normal analytical procedures using the chromatographic conditions detailed below. The values obtained not only include the error in FID signal response but also errors in sample handling and sample loading *i.e.* this method allows the whole analysis process by FID to be assessed. For this reason the value quoted for %RSD for the FID signal response includes the errors inherent in the analytical technique.

Standard conditions:	column temperature	60°C isocratic for 6 mins
	carrier gas	Helium 16psi
	FID	275°C
		40mls.min^{-1} hydrogen
		400mls.min^{-1} air

The lines drawn on the graph (Figure 4.1) are used as guides to identify when the response of the equipment has deviated outside expected limits. The dotted red line represents the average value, where all the points would lie if the equipment was not subject to any error whatsoever. The dotted blue line represents a warning limit which is a deviation of 1.96σ away from the mean (*i.e.* values at the 95% confidence limit). If the results for the test methane sample exceed these limits the analysis system is not working to its optimum. These limits act as a warning and the only action taken when the results exceeded these limits was to perform more regular standard checks to show the value equated to an outlier as opposed to a trend in loss of performance.

If the value obtained for the standard monitoring lies outside the hard black line limit (>3.09σ deviation from the mean *i.e.* values outside the 99.9% confidence limit) or a

series of consecutive results (more than two) consistently lies outside the blue dotted line limit, follow up action has to be taken. The action may range from routine cleaning of the detector to a re-evaluation of any step in the analysis process. The important aspect of this monitoring is to always follow up any result outwith the normal and to find the cause and eliminate the problem. Simply re-analysing the standard to obtain an acceptable result is not acceptable.

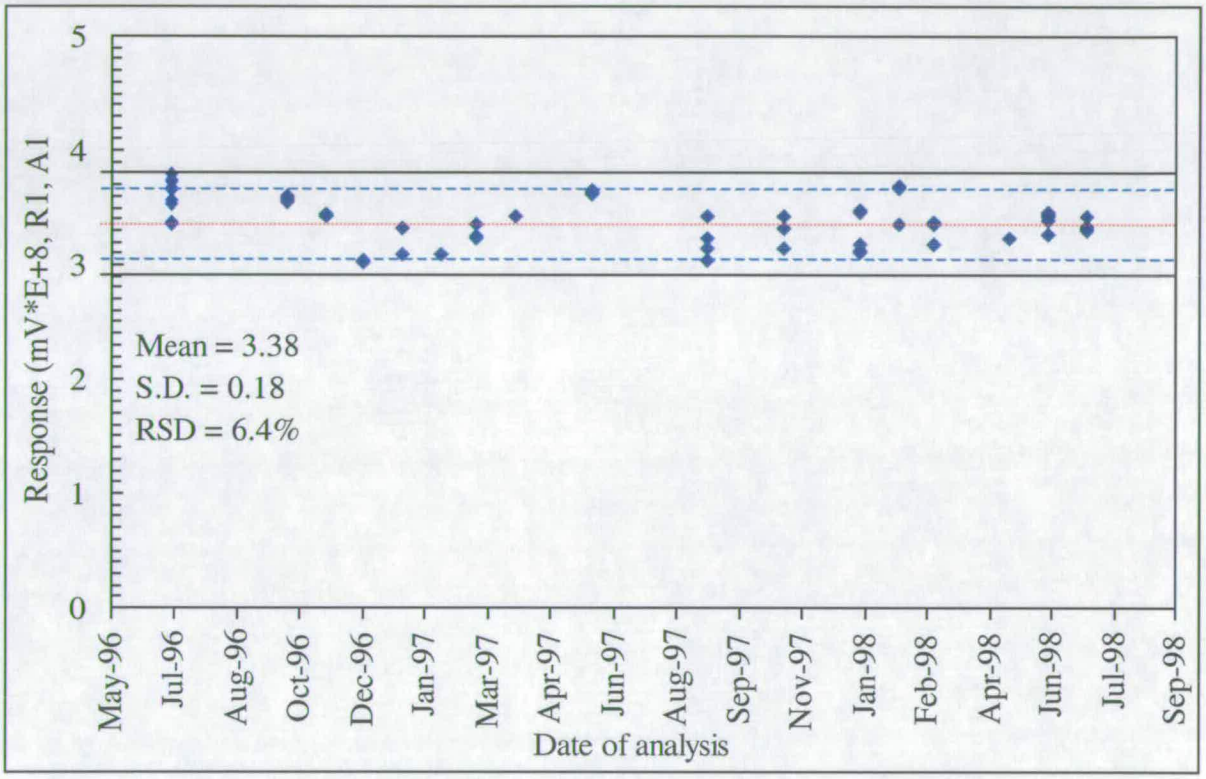


Figure 4.1 *FID signal response to analysis of 50µl of methane standard over the time course of the PhD work (quoted values calculated from accepted monitoring data results).*

4.1.2.1 Examples of problems highlighted with FID monitoring

Example 1 The results for the methane standard were highly erratic with values both above and below the outer monitoring limit. The results appeared to follow no set pattern. On occasion no baseline signal was visible on the integrator.

Inspection of the electrical cable leading from the FID assembly to the circuit board in the GC showed there was a loose connection which intermittently shorted out the signal from the FID to the circuit board.

The cable was replaced and the monitoring results returned to within the expected values.

Example 2 The results for the methane standard began to slowly drift downwards through the warning limit and then towards the outer limit.

As the downward drift in monitoring response indicated a progressive loss of sensitivity the problem appeared to be the detector becoming contaminated.

The FID assembly was stripped down to its component parts and cleaned by sonication in acetone and then in water. The assembly was dried in an oven at 70°C overnight. The FID was re-assembled and the monitoring results were shown to be within the acceptable limits.

4.1.3 Monitoring of ECD performance

As with the FID the ECD signal output was monitored to avoid the possibility that changes in by-product peak heights occurred because of equipment error and not due to differences in samples. The monitoring was performed using an external standard of CFC 115 (chloropentafluoroethane). Methane was not used as it has a reduced response to detection by ECD. CFC 115 was chosen as the standard as the ECD exhibits high sensitivity to this chlorine containing compound and it has a short chromatographic run time (approximately 8mins). 50µl of a diluted sample of CFC 115 (1.279nmoles) was analysed as the response for pure CFC 115 was greater than the upper limit of the ECD. The diluted sample was prepared by diluting 5mls CFC 115 in one bulb volume of helium at 10psi using the dilution system (hence the reported equipment monitoring begins in Aug 96 when this apparatus was built). The diluted CFC 115 standard was analysed under a set of standard conditions every time the ECD was initially switched on and at intervals throughout the time the ECD was being used (approximately once a week).

The monitoring results are displayed in Appendix F with a selection of these results being shown as a Shewhart chart in Figure 4.2. As with the FID monitoring the

%RSD refers to the analysis technique as a whole and not to the variation in the ECD alone. The RSD value of 7.3% shows that the ECD maintained a constant level of response throughout the time employed to analyse samples of HFC 134a. Acceptance limits were set at the same levels as those specified in Section 4.1.2.

Standard conditions:	column temperature	60°C isocratic for 10 mins
	carrier gas	Helium 16psi
	ECD	350°C
		40mls.min^{-1} nitrogen

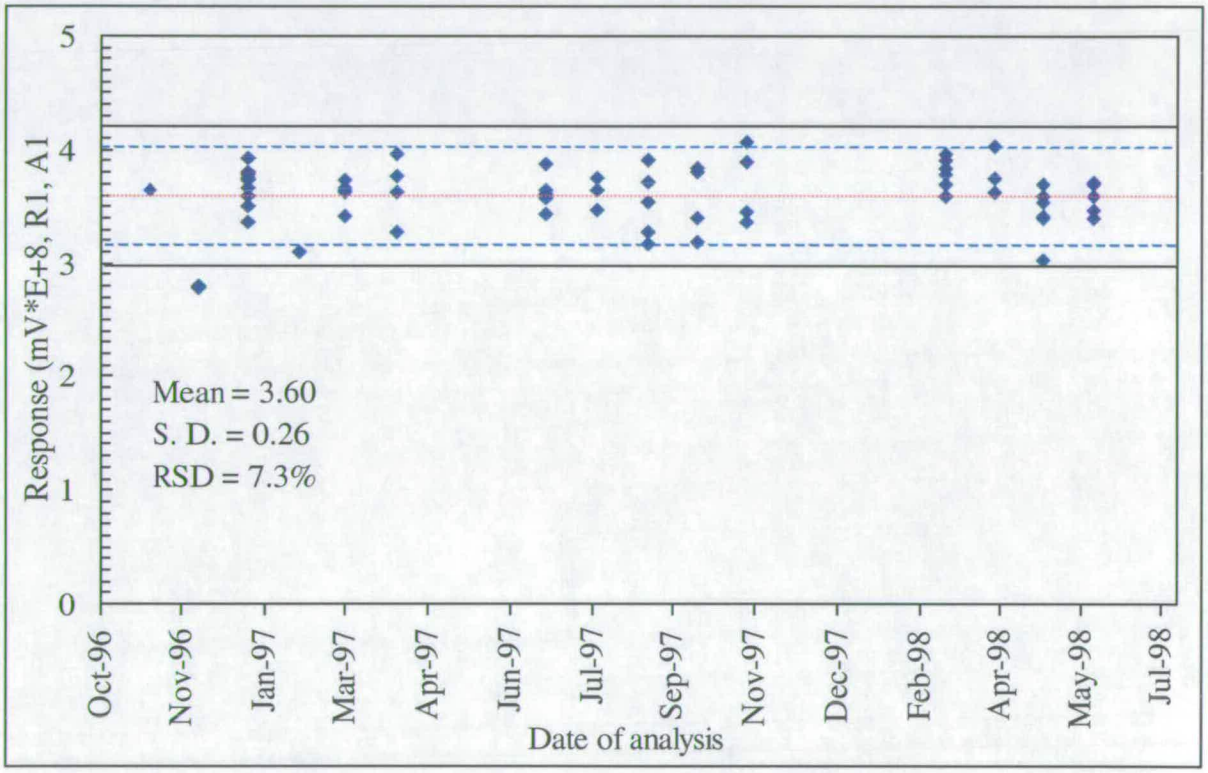


Figure 4.2 *ECD signal response to analysis of 50µl of diluted CFC 115 standard over the time course of the PhD work (quoted values calculated from accepted monitoring data results).*

4.1.3.1 Example of problem highlighted with ECD monitoring

Example 1 The signal for the monitoring result rose over a short period of time towards the upper monitoring limit.

Inspection of the baseline response showed higher levels of noise than were expected (>0.005mV).

The gas filters used to produce ultra-pure nitrogen were inspected and the oxygen filter life expectancy indicator showed the filter was beginning to become saturated. The filter was changed and the monitoring results returned to normal levels.

4.1.4 Linear response range of FID, ECD and MS

The linear range, limit of detection and detector saturation were ascertained for both the FID, ECD and MS by analysing various amounts of standard compounds. The ideal situation when performing analyses which will use peak responses in a quantitative context is to produce calibration lines for all the components detected. That is, analyse all the different components at a range of concentrations from the limit of detection to the point at which saturation of the detector occurs with a range of concentrations at regular intervals. In reality, however, time constraints, availability of authentic standards and the prior knowledge of the required standards will limit the number of linearity graphs possible. It is therefore necessary to take a representative selection of standards and analyse these over their detectable ranges. Careful choice of standards should provide a representative set of calibration curves which can be used for comparison with all the other components detected. From these calibration curves, conclusions on general limits of detection and linearities of response can be made.

The compounds used to assess the linearity of the detectors used in this work were standards of CFC 115 and HFC 134a. These compounds covered the main classes of compounds expected to be encountered in the analysis work performed in the PhD.

4.1.4.1 Linearity of FID signal.

A series of repeat analyses were performed measuring the FID response to various amounts of HFC 134a. 100µl injections of different dilutions of HFC 134a were made starting from the pure gas and then serially diluting the gas until the signal response fell below the limit of detection (see Section 3.4.1). All repeat analyses were performed under the following standard conditions:

Chromatographic conditions	100µl sample loop
	16psi helium carrier gas
	120°C column temperature
FID conditions	Detector temperature 275°C
	40mls.min^{-1} hydrogen 400mls.min^{-1} air

The results for the repeat analyses of the different amounts of HFC 134a are displayed in Figures 4.3a and 4.3b. Figure 4.3a shows the response measured as peak heights and Figure 4.3b describes the response as peak area counts. Both graphs show that the FID had a high degree of linearity ($R^2 = 0.99$) over the concentration range measured (1pg-600µg)^3. The scales of the graphs are in log units to display all the data over the concentration/response range studied. There were no major differences in linearities when using either the height or area measurements to describe the detected compound peak.

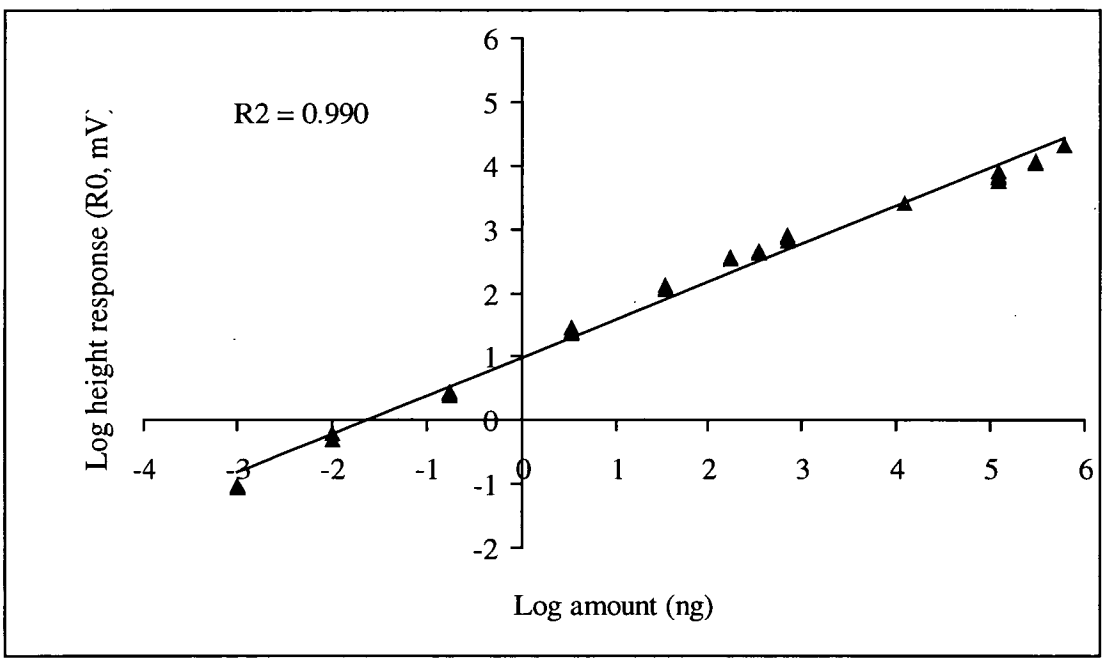


Figure 4.3a *Linearity plot of height response for FID analysis of HFC 134a.*

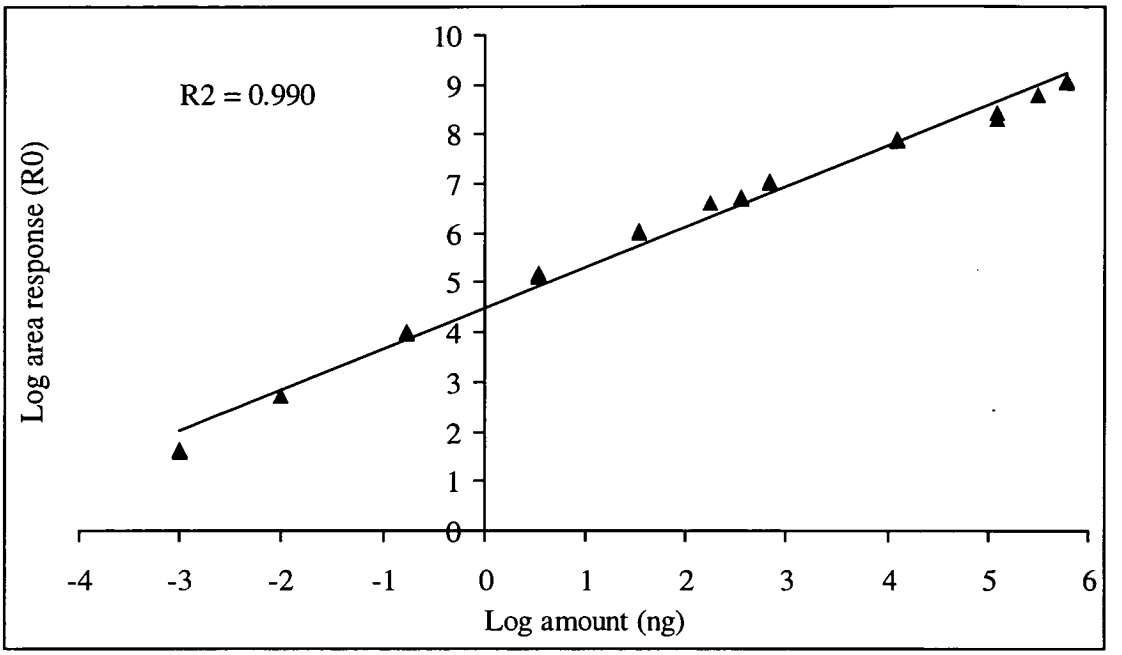


Figure 4.3b *Linearity plot of area response for FID analysis of HFC 134a.*

4.1.4.2 Linearity of ECD signal.

The ECD linear response range was investigated by performing repeat measurements of various amounts of CFC 115. 100µl injections of different dilutions of CFC 115 were made from a starting dilution (equivalent to 688ng) and then serially diluting this sample until the signal response fell below the limit of detection (see Section 3.4.2). All repeat analyses were performed under the following standard conditions:

Chromatographic conditions	100µl sample loop
	16psi helium carrier gas
	60°C column temperature
ECD conditions	Detector temperature 350°C
	40mls.min^{-1} nitrogen

The results for the repeat analyses of the different amounts of CFC 115 are displayed in Figures 4.4a and 4.4b. Figure 4.4a shows the response measured as peak heights and Figure 4.4b describes the response as peak area counts. Both graphs show that the ECD had a high degree of linearity ($R^2 > 0.99$) over the concentration range measured (10fg-688ng)[4]. It should be noted that the direct comparison of the FID and ECD linearities can not be made as the number of points used to calculate R^2 in

the ECD experiment is greater than in the FID experiment[4]. The scales of the graphs are in Log units to display all the data over the concentration/response range studied. There were no major differences in linearities when using either the height or area measurements to describe the detected compound peak.

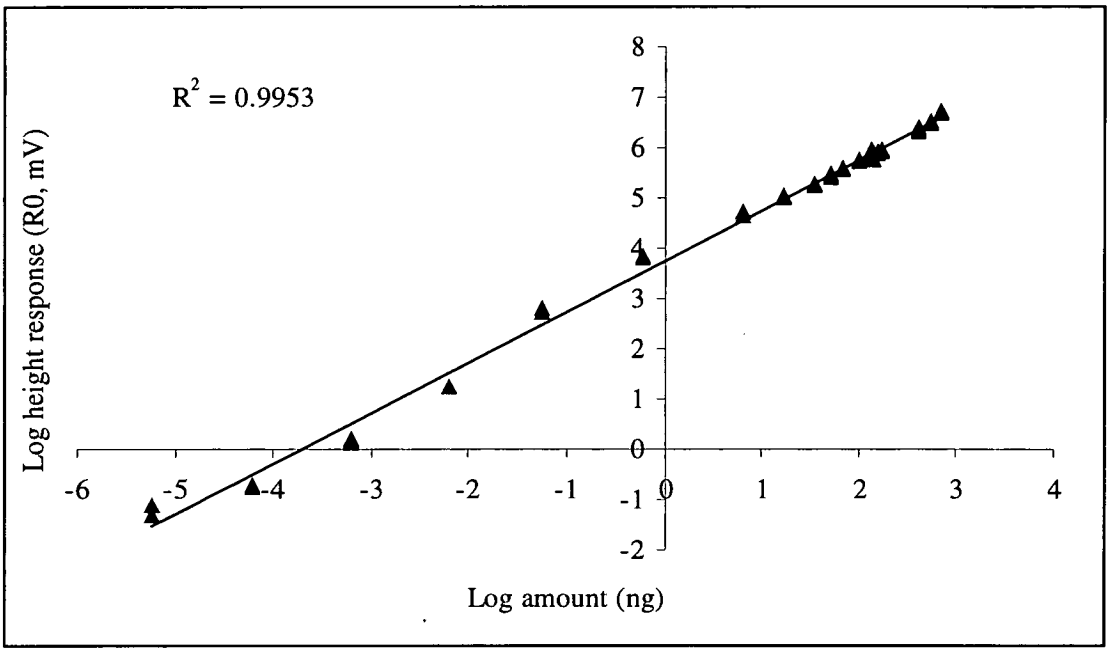


Figure 4.4a *Linearity plot of height response for ECD analysis of CFC 115.*

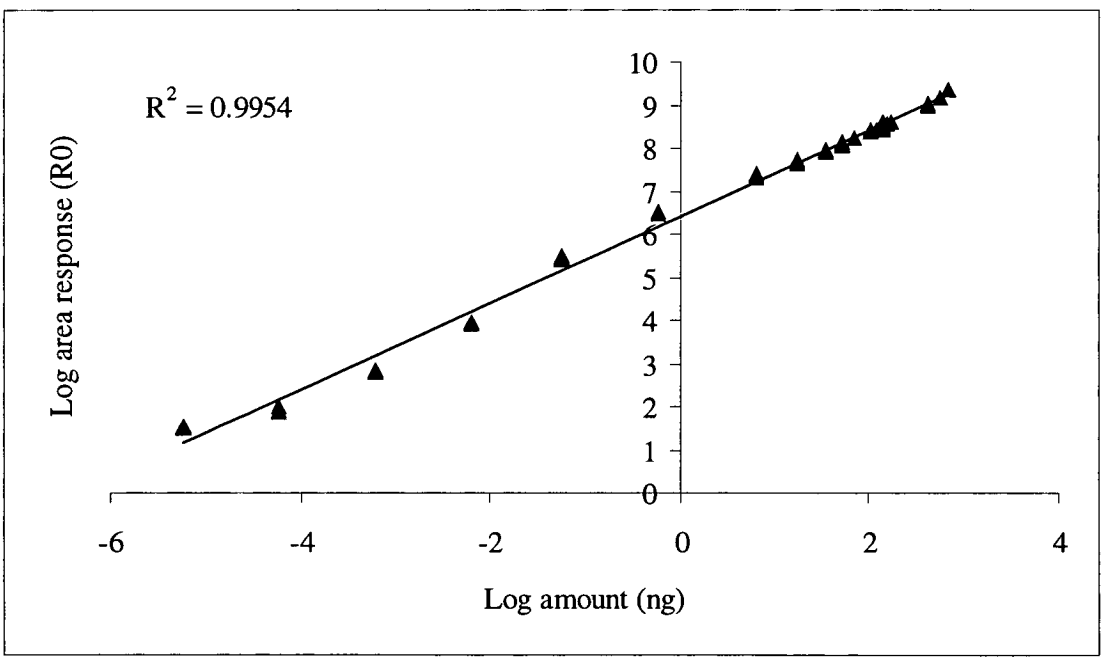


Figure 4.4b *Linearity plot of area response for ECD analysis of CFC 115.*

4.1.4.3 Linearity of Mass spectrometer signal.

A series of repeat injections of various amounts of HFC 134a were made into the GC-MS system. The GC-MS was operated under the standard conditions detailed below. Single ion monitoring was not used to measure the samples since, although it would be more sensitive, it was felt this would not reflect the actual conditions under which samples would be analysed.

GC conditions	100µl sample loop
	16psi helium carrier gas
	100°C column temperature
MS conditions	70eV ion current
	30-120Da scan range (1 scan per second)
	1.5kV electron multiplier

The results for these analyses are shown in Figures 4.5a and 4.5b. Figure 4.5a shows the results obtained from measuring the peak height response for the TIC and Figure 4.5b shows the results when measuring the peak area response for the TIC. Each concentration measurement was performed at least three times. Both methods show that the mass spectrometer response is not linear over the whole concentration range studied. However, both graphs are linear over the lower concentration ranges (<10 and <100µg for the height and area measurements respectively) which are in the region of the amounts of the by-products expected to be seen. It is also apparent from the two graphs that the measurement of peak heights is more susceptible to non-linearity at higher sample concentrations than the measurement of peak area.

The non-linearity of the mass spectrometry response is because of a number of different factors although the main one is saturation of the detector. To overcome the problem, the voltage on the electron multiplier can be lowered. This reduces the overall level of signal and extends the effective range of the instrument. It does however, decrease the level of sensitivity so when analysing mixtures anything present at low concentrations may not be seen. For this reason the mass spectrometer was operated under optimised conditions for sensitivity as the by-products present in

the samples of HFC 134a were at low concentrations. That is, effects of non-linear instrument performance were unlikely to cause deterioration in the quality of the MS data.

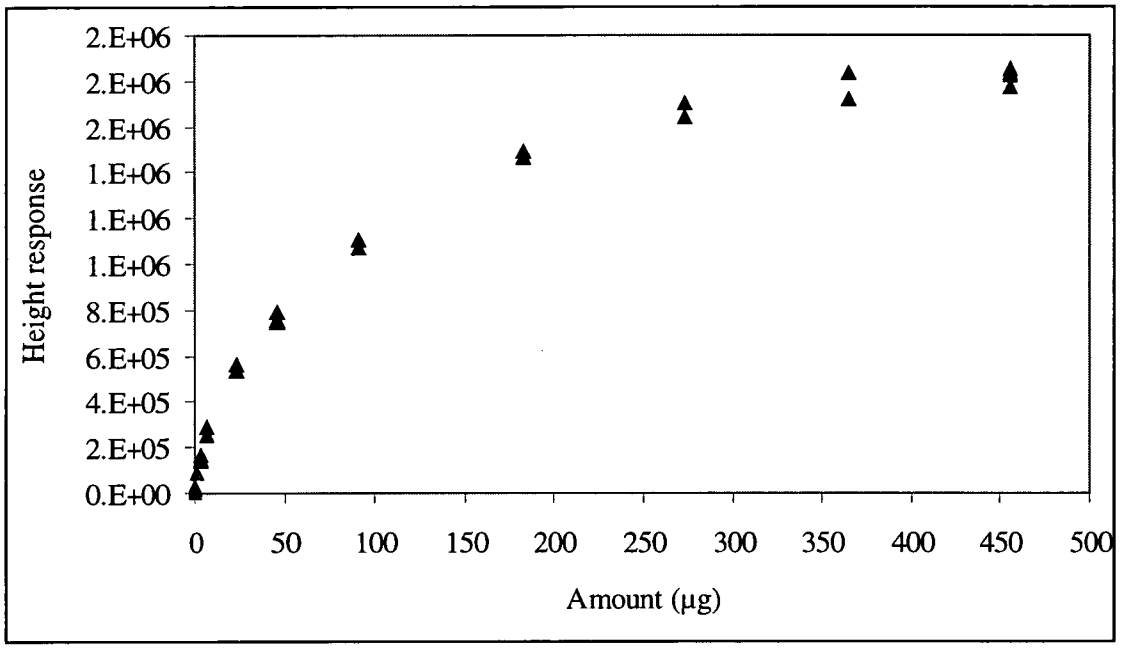


Figure 4.5a *Mass spectrometer peak height response for samples of HFC 134a.*

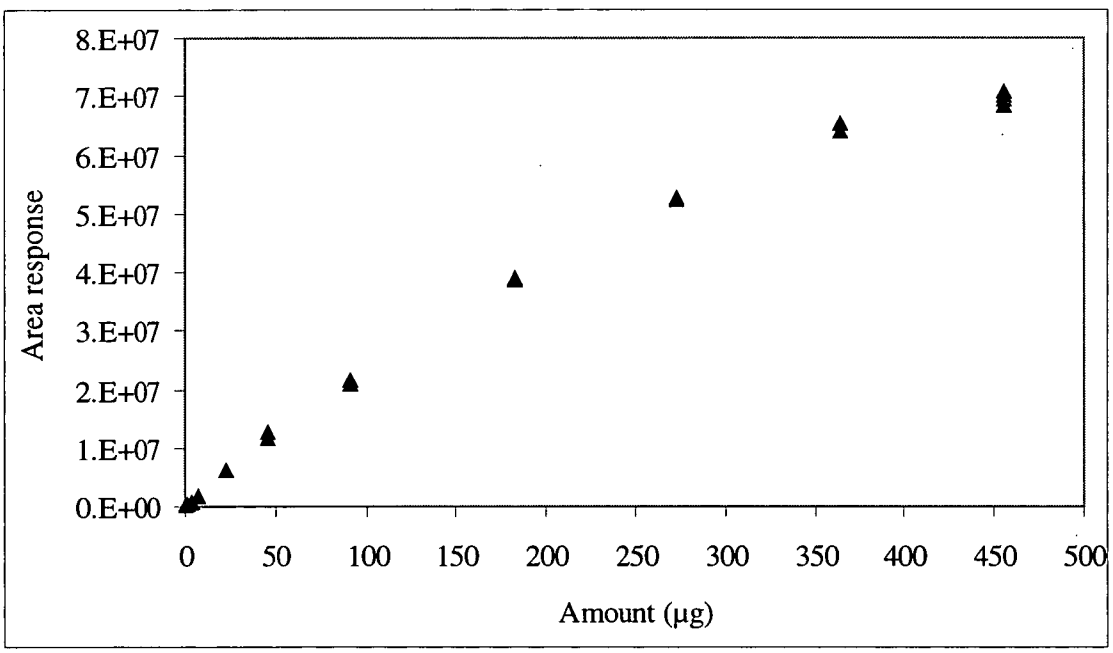


Figure 4.5b *Mass spectrometer peak area response for samples of HFC 134a.*

4.2 Development of separation temperature program

The separation of the complex mixtures of low molecular weight halocarbons requires a multi-step temperature program[5]. The following sections describe the experiments performed to optimise these temperature steps to give the best separation. The best separation can be considered as when the majority of by-products are separated both from each other and from the swamping effect of the large HFC 134a peak. To define the term 'best' is difficult as a number of factors contribute in a variety of manners in producing an ideal chromatographic profile. These factors can sometimes be contradictory; however, in this work chromatographic run time is not a limiting factor and therefore the highest degree of peak resolution is the main consideration. This allows the use of long run times (>30mins) to obtain the maximum amount of component separation.

4.2.1 Initial temperature

Separation of low molecular weight halocarbons using PLOT alumina columns has shown that potential by-products will chromatograph both before and after HFC 134a[6]. The relatively low levels of by-products present (<100ppm) dictates that the HFC 134a samples will be loaded at levels where HFC 134a will overload the column. This will produce a large peak, corresponding to HFC 134a, which will have both a leading edge which swamps peaks chromatographing before HFC 134a and a large amount of peak tailing which will swamp peaks chromatographing after HFC 134a (see Figure 4.6)[7].

To avoid swamping of peaks chromatographing before HFC 134a an isocratic temperature is used to allow these by-products to chromatograph before HFC 134a appears. A series of experiments were performed using a range of isocratic column temperatures (70-40°C) which were held until the leading edge of the HFC 134a peak was seen. Once the leading edge of the peak for HFC 134a was seen the column temperature was increased to 180°C at $10^{\circ}.min^{-1}$ and held for 20mins to chromatograph all other by-products. The results showed that at initial column temperatures greater than 60°C the leading edge of the HFC 134a peak swamps a

number of by-product peaks. At the lowest temperature of 40°C the HFC 134a peak does not chromatograph until >40 mins which increases the chromatographic run time and also affects the peak shape of later eluting by-products. For these reasons a compromise temperature of 50°C was chosen which allows the by-product peaks to be separated from HFC 134a without too long an initial hold time.

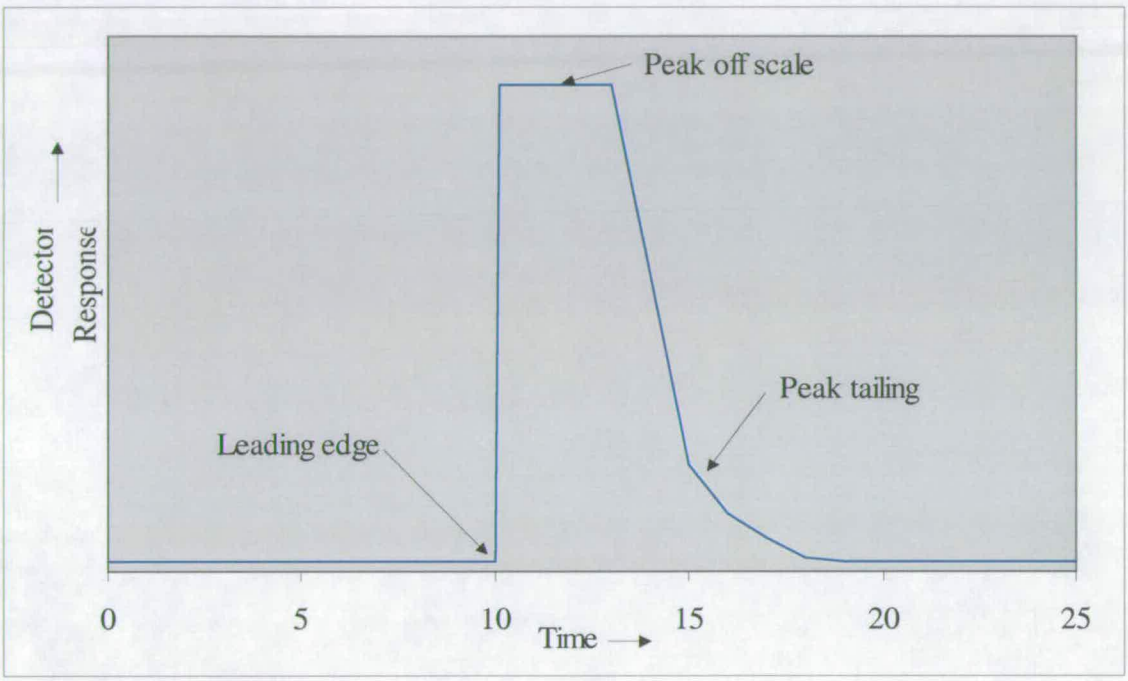


Figure 4.6 *Diagram of overloaded peak.*

4.2.2 Multi-step temperature programming

After the HFC 134a peak begins to chromatograph the temperature can be increased rapidly in order to elute all the HFC 134a. The temperature is increased to a set point and held at this temperature to allow the complete elution of HFC 134a and its tailing peak before the next by-product elutes. A series of temperatures were investigated to find the one which separated the next chromatographing by-product from the swamping effect of HFC 134a. The results showed that holding at 100°C separates the by-products appearing after HFC 134a from each other and from the tailing effect of the overloaded HFC 134a.

A further temperature step is added to elute the components most retained by the column (*e.g.* HCFC 123). This involves raising the temperature to 180°C (20°C

below the column maximum operating temperature) at 10°C.min^{-1} and holding for 5 mins. The rate of temperature change increases the baseline drift and noise and therefore rates >10°C.min^{-1} cause serious detriment to the quality of the chromatographic profile.

4.2.3 Chromatographic temperature conditions for analysis of HFC 134a

The multi-step temperature separation method used for all the analyses involving samples of HFC 134a (unless otherwise stated) is as follows:

50°C for 33mins, increase at 20°.min^{-1} to 100°C
100°C for 20 mins, increase at 10°.min^{-1} to180°C
180°C for 5 mins.

4.3 Chromatographic profiles obtained for FID and ECDanalysis

The HFC 134a samples of known origin were analysed by GC-FID and GC-ECD using the standard temperature program detailed in Section 4.2.3. Each sample was chromatographically profiled in replicate (n=5) by each detection method. The results for these analyses are displayed in Appendices G and H. An examination of a typical set of profiling results for one of the samples of HFC 134a will serve to illustrate the differences in the data sets obtained by the two methods of detection and the way in which the data are handled and interpreted.

4.3.1 FID results for sample of HFC 134a produced by ICI in the UK

A typical profile for one of the replicate analyses of the HFC 134a sample from ICI UK is displayed in Figure 4.7. The component with the large peak height labelled 'HFC 134a peak' represents HFC 134a and shows the difference in concentration levels between it and the by-products. The chromatogram can be considered in two parts, the area before HFC 134a elutes and the area after the elution of HFC 134a.

The area before HFC 134a elutes shows the presence of seven major peaks (the peak just before HFC 134a is not shown due to limits of the printer) which have

signal/noise (s/n) ratios >3:1 (see Section 1.8.3). However, there are two peaks (not apparent in the figure because of resolution effects) whose s/n ratios are <3:1 yet can be clearly identified in all the replicate analyses. As these peaks appear in each chromatogram with same retention time they are representative of by-products. Therefore, although they do not meet the s/n ratio criteria they are accepted as peaks, and are measured and used in profiling the sample.

The tailing edge of the HFC 134a peak has a number of apparent peaks present. These are, however, artefacts caused by the overloading of the column and detector. The region immediately after tailing stops contains two peaks (labelled Peak 8 and 9) corresponding to by-products present in the sample. A number of smaller peaks (peak heights with s/n ratios in the region of 3:1) were seen in some of the replicate chromatograms but they did not appear with constant retention times and peak heights and were therefore not considered as real peaks. The increase in temperature to 180°C leads to an increase in background noise and although a number of peaks are seen in this region they do not have consistent peak heights or retention times. This suggests that these are not sample related and may represent contamination from the system (*e.g.* column bleed) due to the high temperatures involved. This area of the chromatogram is still monitored as late running peaks (*e.g.* HFC 123) have been shown to have chromatographic run times in this region.

The retention times and peak heights were measured for each peak seen in the chromatographic traces. The retention time relative to the leading edge of the HFC 134a peak was calculated for each peak (Appendix G). These values were taken for subsequent use in the chemometric analysis of the data.

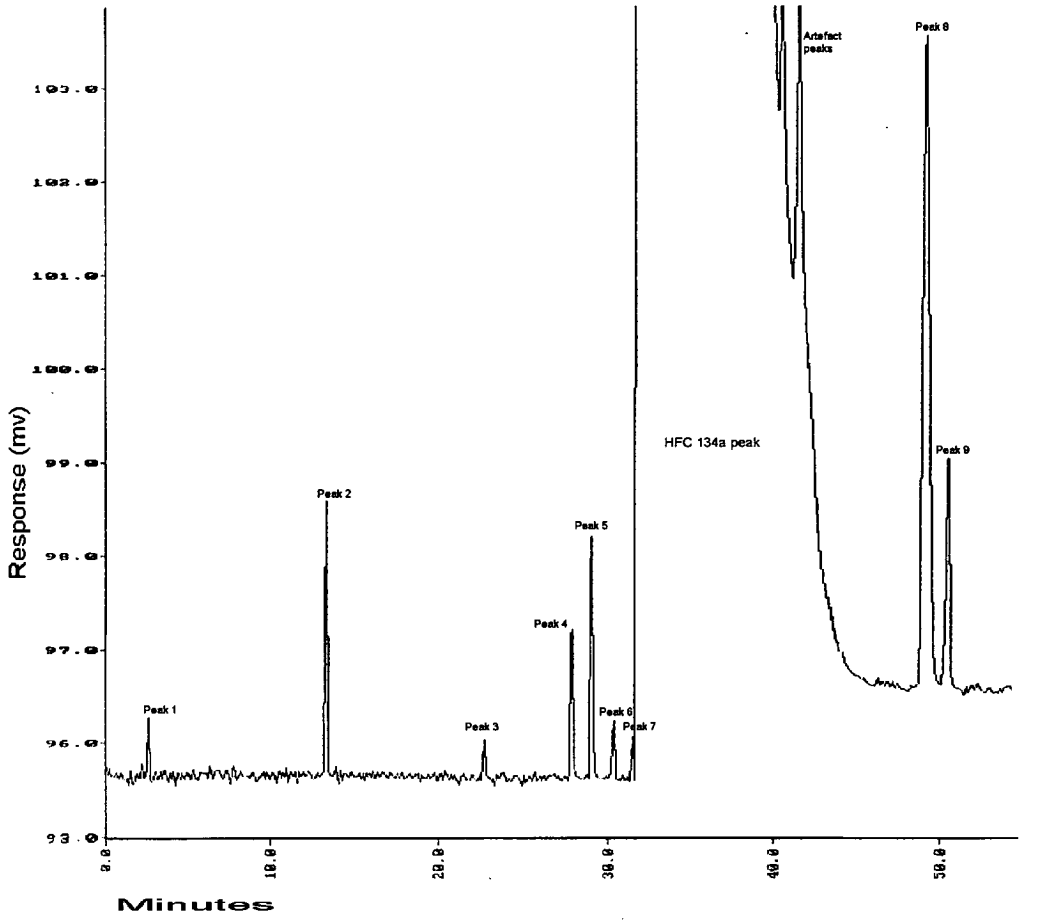


Figure 4.7 *FID chromatographic profile for a replicate analysis of an HFC 134a sample produced by ICI UK.*

4.3.2 ECD results for profiling ICI UK sample of HFC 134a

A typical profile of one of the replicate analyses of the HFC 134a sample from ICI UK is displayed in Figure 4.8 (all ECD profiles are detailed in Appendix H). Comparing the FID profile in Figure 4.7 and the ECD profile it is immediately apparent that the two detection techniques produce vastly different profiles. In the ECD profile the peak representing HFC 134a has reduced dramatically in height; the ECD has low relative sensitivity to this compound, as it does not contain chlorine. The strange peak shape is due to the temperature increase occurring as the peak elutes causing baseline drift into the spike seen at approximately 43 minutes.

The peaks eluting before HFC 134a also have dramatically altered peak heights when compared with the FID profile. Not all the peaks seen in the FID chromatogram are

present in the ECD profile and a number of new peaks are seen in the ECD profile. Increases in the peak heights for peaks labelled 4, 5, 6 and 7 would suggest that these compounds are chlorine containing by-products and therefore show increased levels of sensitivity when using ECD.

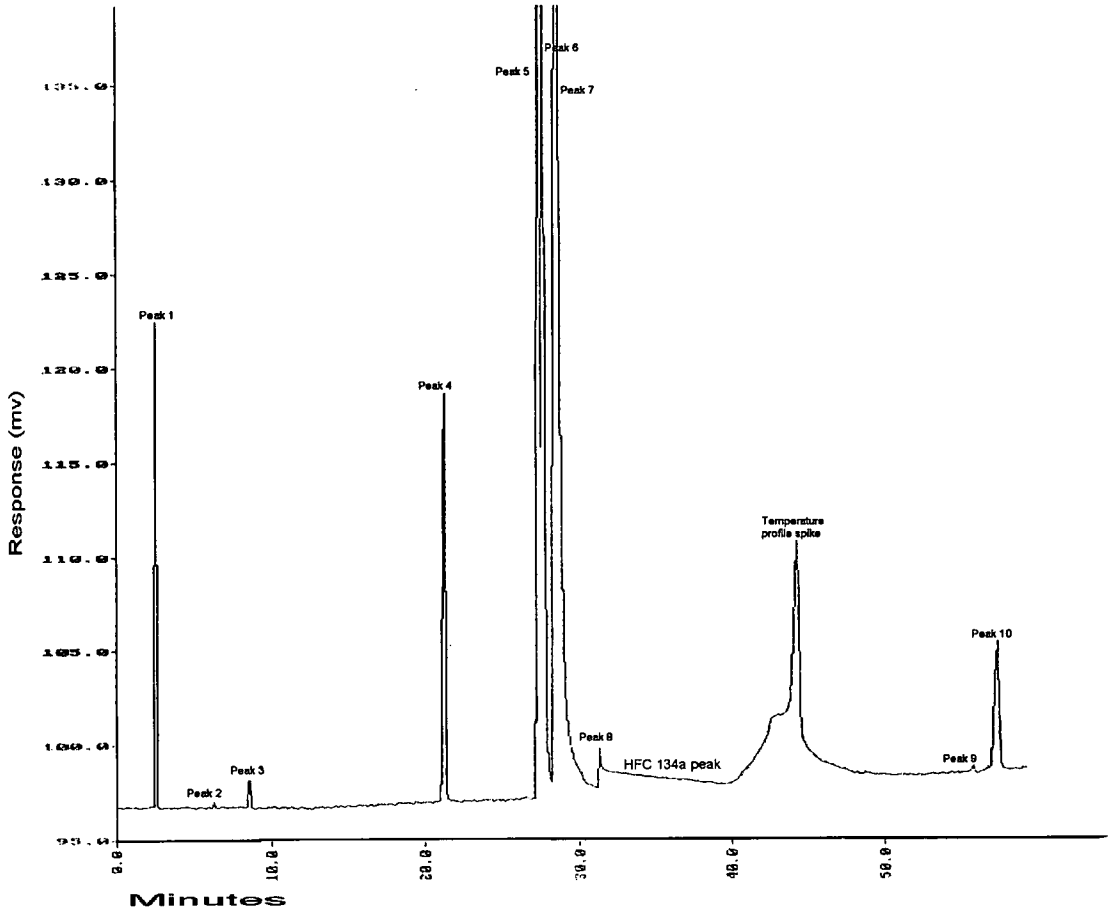


Figure 4.8 *ECD chromatographic profile for a replicate analysis of an HFC 134a sample produced by ICI UK*

The ECD chromatogram contains a number of peaks after the HFC 134a peak. Again, the heights of these peaks differ from those seen in FID chromatogram which may indicate these are different components to those seen in the FID or more probably be because of the different levels of sensitivity due to the presence of chlorine containing molecules.

In the later region of the chromatogram (>60 mins) there are a number of large peaks. As with the FID chromatogram these peaks showed large variations between replicates and were not seen in all samples of HFC 134a analysed. This suggests that

these peaks are due to either contamination from the system (*e.g.* column bleed) or from outside the system (*e.g.* trace levels of solvents in the laboratory atmosphere). The size of these peaks are relatively larger in the ECD trace than in the FID trace because of the higher sensitivity of the ECD to potential contaminants (*e.g.* chlorinated solvents).

4.4 Identification of by-products by GC-MS

The following section details how the mass spectra are interpreted by using the analysis of an HFC 134a sample produced by ICI UK as an example. The full interpretation and *m/z* values of the ions seen in all the samples analysed are detailed in Appendix I. Because of the limited amount of access time to mass spectrometer equipment a reduced chromatographic run time was employed. The data produced by the mass spectrometer allows co-eluting peaks to be distinguished on the basis of the ions present. This means the temperature program for the GC-MS work can be shortened which leads to co-elution of peaks (in a few cases) which can still be distinguished and identified. The temperature gradient employed was as follows:

50°C for 5.5 mins, increase at $4^{\circ}.min^{-1}$ to 80°C

Increase at $10^{\circ}.min^{-1}$ to100°C hold for 4 minutes

Increase at $20^{\circ}.min^{-1}$ to180°C hold for 5 mins.

Figure 4.9a shows the TIC chromatogram for a replicate analysis of an HFC 134a sample produced by ICI UK and Figure 4.9b is an enlargement of the front portion of this trace. Comparing this chromatogram to the FID and ECD trace of the same sample shows that the profile is most similar to the FID trace. This is as expected as the mass spectrometer does not have particularly high sensitivity to any of the compounds studied[8]. The MS profile shows a number of peaks before and after the large peak representing HFC 134a.

The mass spectra, which are summed to give the TIC, can be viewed individually in order to see what ions are present at any point in the chromatographic trace[9]. The mass spectrum at any point in the chromatogram can be compared with previously

analysed authentic standards or interpreted directly by the operator. In the majority of cases detailed in this work the identification was made on the basis of previously analysed standards. Identification of components based wholly on the interpretation of mass spectra may lead to errors as discrimination between isomers is not possible unless a spectrum from a previously analysed standard is available for comparison.

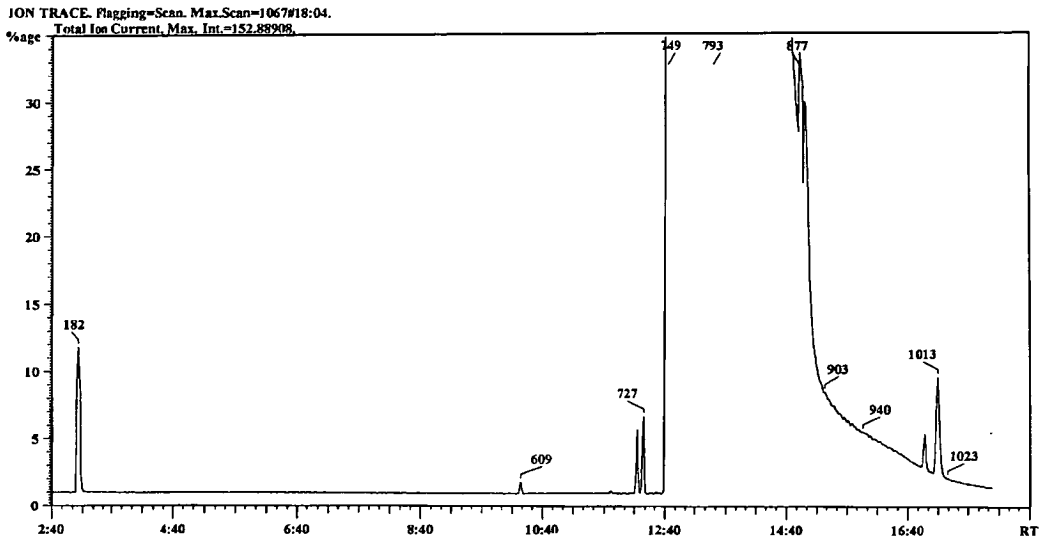


Figure 4.9a *TIC chromatographic profile obtained from the mass spectral analysis of an HFC 134a sample produced by ICI UK.*

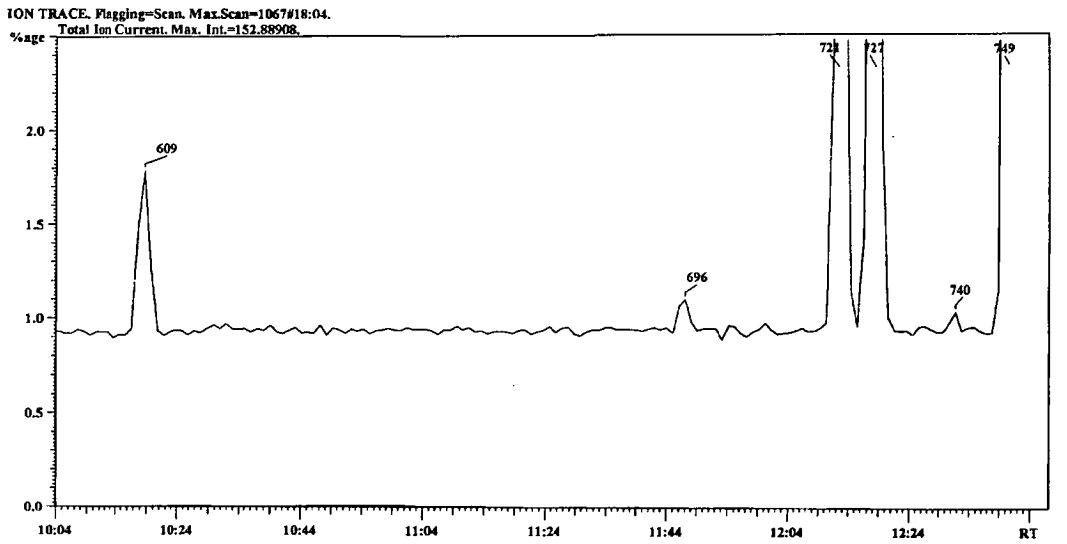


Figure 4.9b *Enlargement of Figure 4.9a of region before HFC 134a peak.*

The TIC is a representation of all the ions present over the mass range scanned. The TIC can be expanded to show the mass spectra at any point in the chromatographic

trace. By viewing the mass spectrum representing the different peaks seen in the TIC trace the identity of the components can be elucidated. Single ion chromatograms can also be displayed in order to check that the presence of an ion with a specific *m/z* value in a spectrum is not due to a background effect. The mass spectra representing the peaks seen in the TIC for the analysis of HFC 134a are displayed in Figures 4.10-4.17.

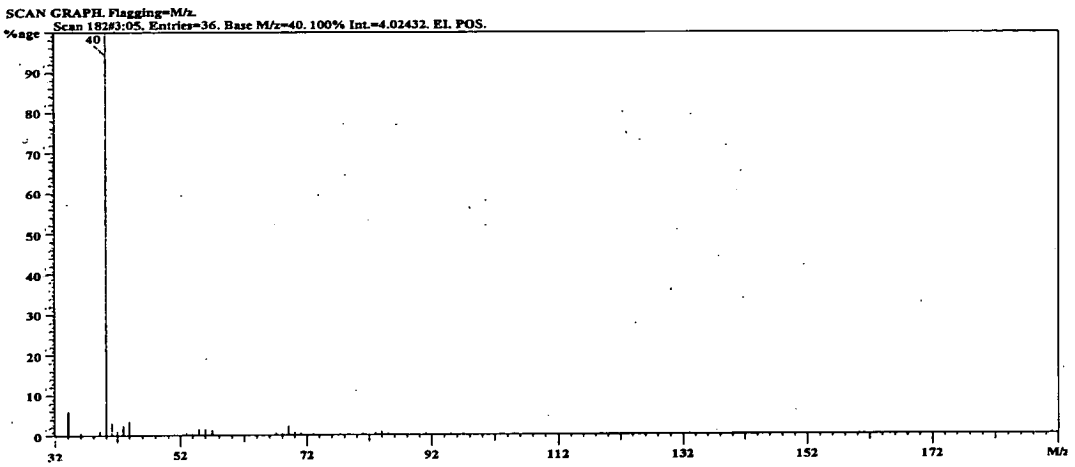


Figure 4.10 *Mass spectrum for peak at scan number 182.*

Figure 4.10 shows the mass spectrum for the first peak in the TIC at scan number 182. The mass spectrum corresponds to air and represents the retention time corresponding to the dead volume of the column.

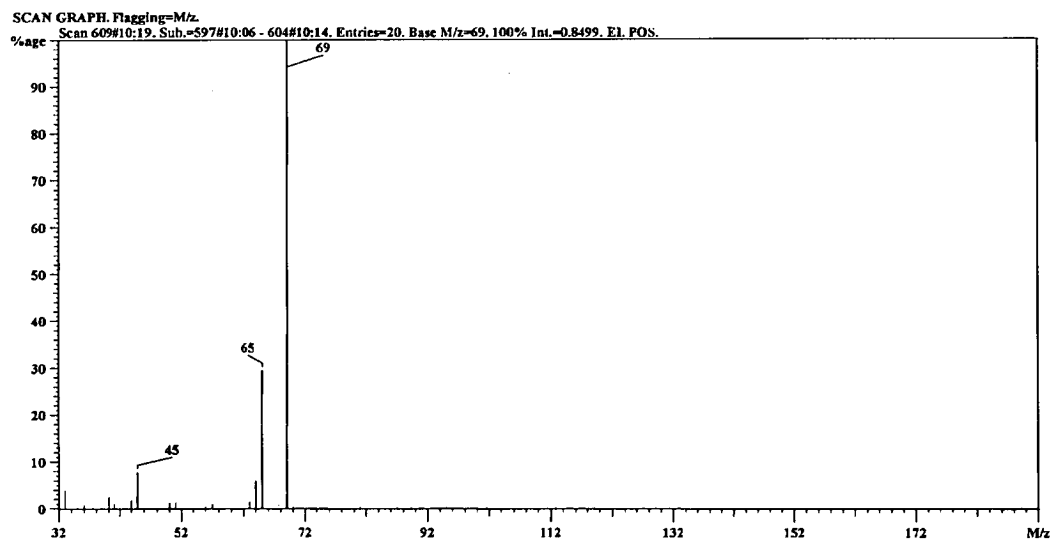


Figure 4.11 *Mass spectrum for peak at scan number 609.*

Figure 4.11 shows the mass spectrum for the peak at scan number 609 (retention time 10:19). A library search showed the mass spectrum to correspond to HFC 143a (CF_3CH_3) with the characteristic ions with m/z values corresponding to CF_3 *m/z* = 69, CH_3CF_2 *m/z* = 65 and CH_2CF *m/z* = 45. No molecular ion with a *m/z* value of 84 was seen for this compound.

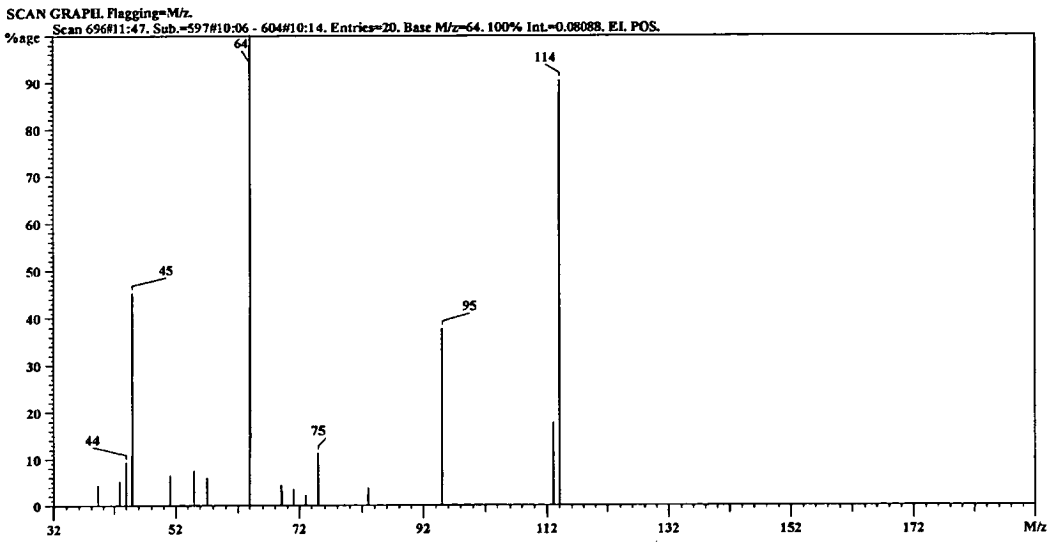


Figure 4.12 *Mass spectrum for peak at scan number 696.*

Figure 4.12 shows the mass spectrum for the peak at 696 (RT 11:47mins). Assignment of ions in the spectrum was as follows, $C_3H_2F_4$ *m/z* = 114, $C_3H_2F_3$ *m/z* = 95, C_3HF_2 *m/z* = 75, $C_2H_2F_2$ *m/z* = 64 and C_2H_2F *m/z* = 45. Comparing this spectrum with previously analysed standards showed the closest match to HFC 1234yf (CH_3CHFCF_3); however, the two spectra differ in that the analysed sample does not contain an ion at *m/z* = 69 (CF_3). This suggests that the by-product in the sample is a different isomer of HFC 1234yf.

Figure 4.13 displays the mass spectrum for the peak at 721 (RT 12:12mins). Assignment of ions was as follows, $C_2F_3Cl_2$ *m/z* = 151, C_2F_4Cl *m/z* = 135, C_2F_4H or CFCl *m/z* = 101, CF_2Cl *m/z* = 85 and CF_2 *m/z* = 50. The assignment of the chlorine containing species is made easier by the isotopic ratio seen in the mass spectra of 3:1 between ^{35}Cl and ^{37}Cl (*e.g.* the ions at *m/z* 135 and 137). Comparison of this

spectrum with a library of spectra showed the compound to be CFC 114 (CF_2ClCF_2Cl).

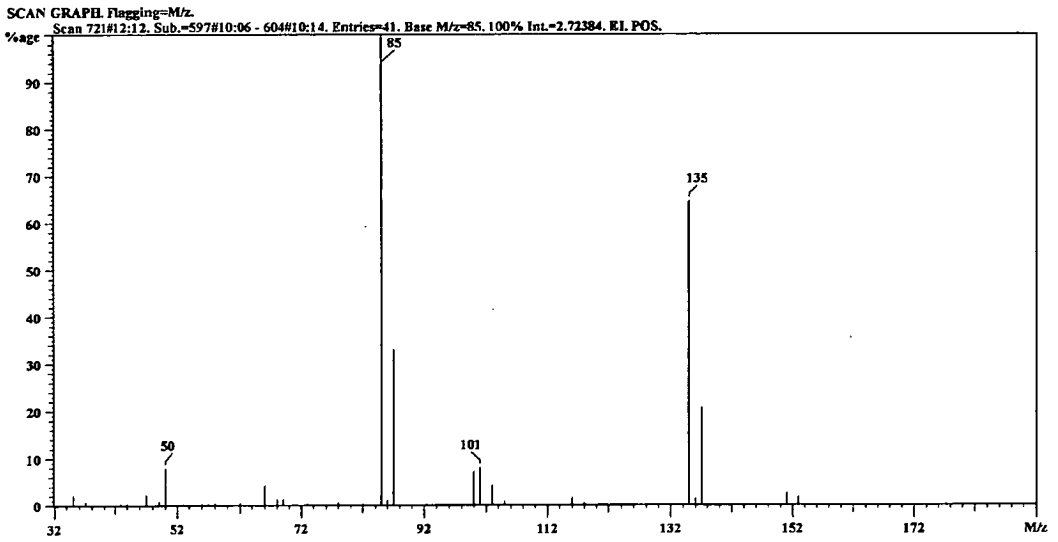


Figure 4.13 *Mass spectrum for peak at scan number 721.*

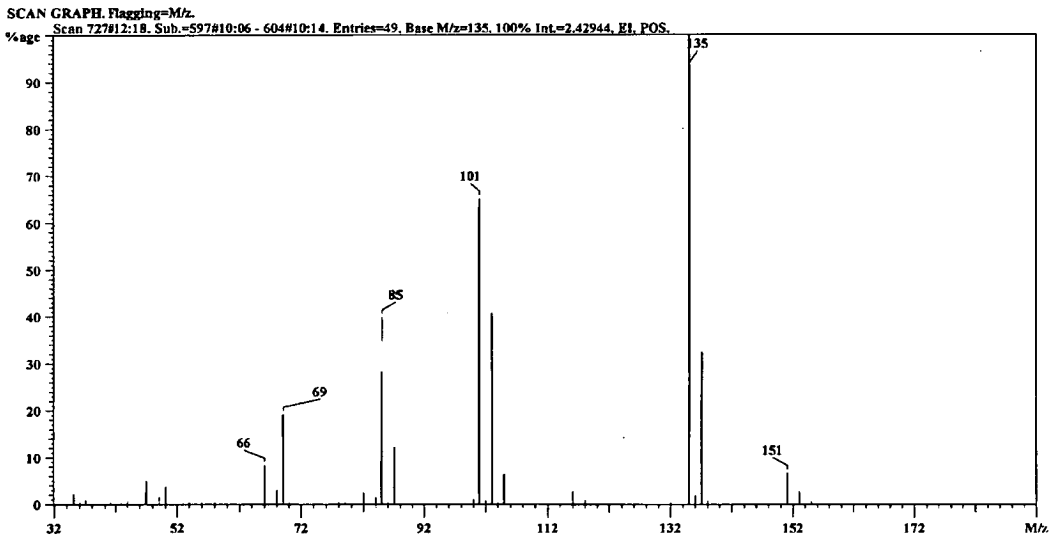


Figure 4.14 *Mass spectrum for peak at scan number 727.*

The spectrum in Figure 4.14 is taken at scan 727 (RT 12:18mins). Assignment of ions was as follows, $C_2F_3Cl_2$ *m/z* =151, C_2F_4Cl *m/z* = 135, $CFCl_2$ *m/z* = 101, CF_2Cl *m/z* = 85, CF_3 *m/z* = 69 and CFCl *m/z* = 66. The assignment of the single chlorine containing species was performed as in spectrum 721 and the assignment of two chlorine atom containing species was aided by the isotopic ratio pattern of 9:6:1.

This isotopic ratio pattern of a molecule containing two chlorine atoms occurs because of the presence of either two ^{35}Cl isotopes, one isotope of ^{35}Cl and one of ^{37}Cl, or two ^{37}Cl isotopes. A library search showed this spectrum to match that of CFC 114a (CF_3CFCl_2).

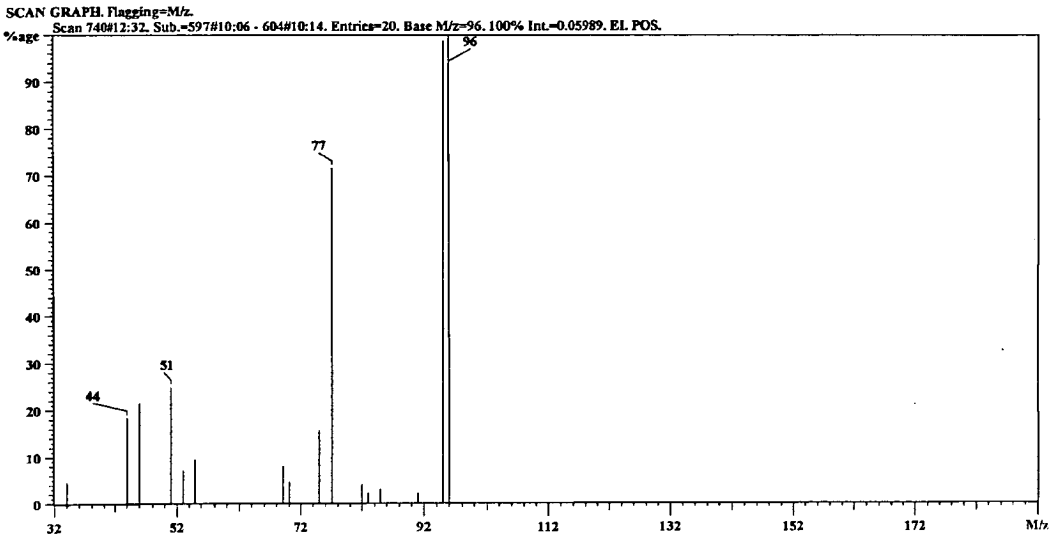


Figure 4.15 *Mass spectrum for peak at scan number 740.*

The spectrum at scan 740 (12:32mins) is displayed in Figure 4.15. The ions were assigned as follows, $C_3H_3F_3$ *m/z* = 96, $C_3H_2F_3$ *m/z* = 95, $C_3H_3F_2$ *m/z* = 77, CHF_2 *m/z* = 51, C_2H_3F *m/z* = 46 and C_2HF *m/z* = 44. Comparison of this component with library spectra showed similarity with HFC 1243zf (CF_3CHCH_2); however, the relative abundance's of a number of the peaks are different suggesting that the component analysed is a different isomer of HFC 1243.

The mass spectrum for the peak at scan number 747 (RT 12:39mins) is shown in Figure 4.16. The intensity of this peak is low and therefore the spectrum obtained is of low quality. The assignment of ions is as follows, $C_2H_3F_2$ *m/z* = 65, CHF_2 *m/z* = 51, C_2H_2F *m/z* = 45 and C_2HF *m/z* = 44. Comparison with the available standards shows close similarity between the spectra seen in the sample and that of HFC 152a. However, relative peak intensities suggest that the compound is not HFC 152a (CF_2HCH_3) but more likely HFC 152 (CFH_2CFH_2).

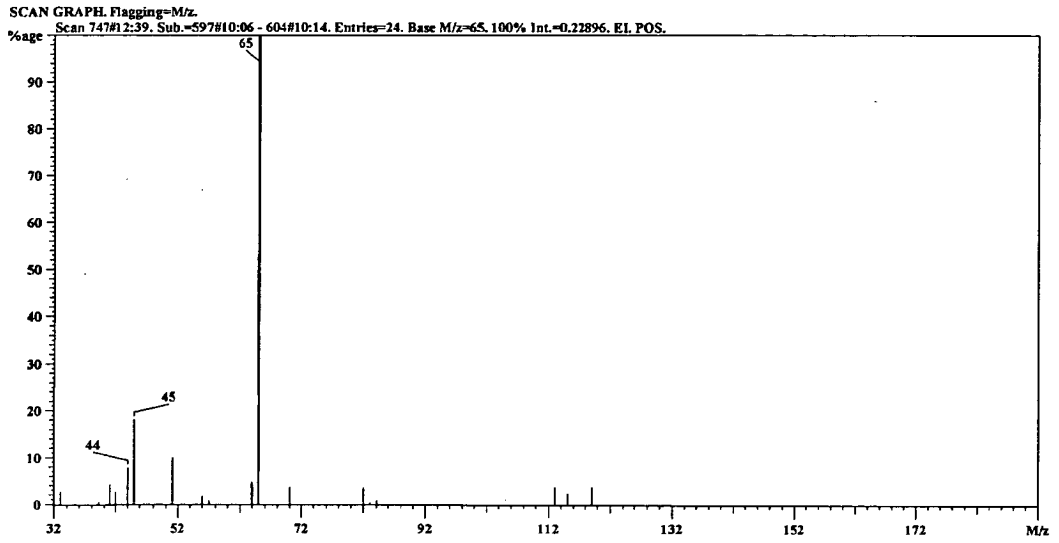


Figure 4.16 *Mass spectrum for peak at scan number 747.*

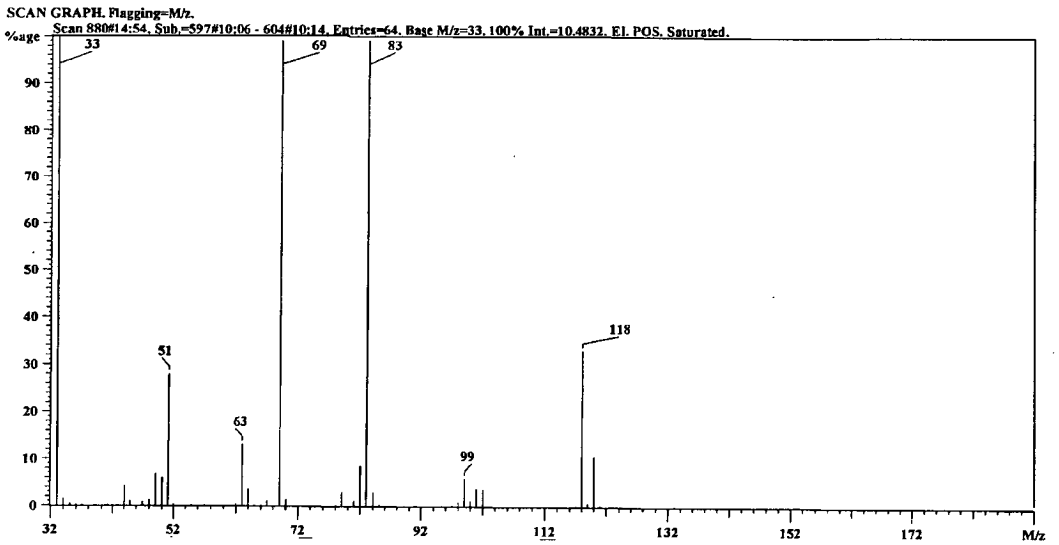


Figure 4.17 *Mass spectrum for peak at scan number 880.*

The mass spectrum for the major peak (Figure 4.17) shows some distortion of ions because of the mass spectrometer detection system being overloaded. The characteristic ions of C_2HF_4 *m/z* = 101, C_2H_2F *m/z* = 83, CF_3 *m/z* = 69, C_2HF_2 *m/z* = 63, CHF_2 *m/z* = 51 and CH_2F *m/z* = 33 are seen, confirming this peak is HFC 134a.

In the TIC the tailing region of the HFC 134a peak contains a couple of peaks. On examination of these spectra the ions seen are the same as those in the normal HFC 134a spectrum with the inclusion of $C_2H_2F_3Cl$ *m/z* = 118. Altering the chromatographic conditions did not affect the retention of these peaks which suggests that they appear as a consequence of the overloading of the HFC 134a sample and are not distinct components. The presence of these peaks may be because of the catalytic effect of the column causing the substitution of Cl into HFC 134a[10]. This extra component/ion may also explain the presence of the spike/peak seen in the ECD chromatographic trace (Figure 4.8 Section 4.3.2). The exact origin of the Cl atom is unclear. It may be from the dehalogenation of another component present in the sample or be from endogenous sources within the sample of HFC 134a. As these peaks for the chlorinated HFC 134a are seen in all the samples of HFC 134a analysed it is doubtful that they are due to the presence of a specific component and more likely that they arise from the effect of such high levels of HFC 134a present in the column at the elevated temperatures used in the separation method.

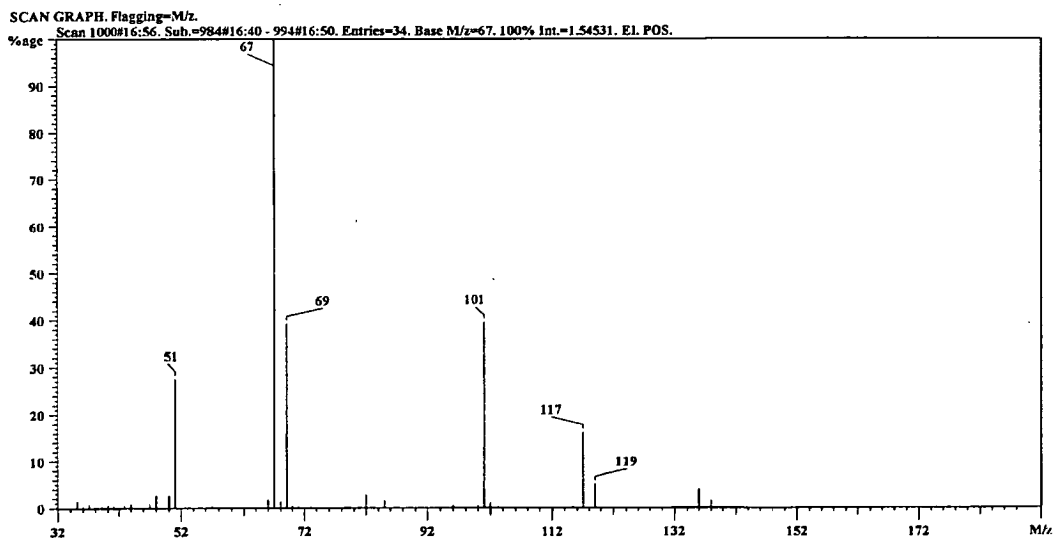


Figure 4.18 *Mass spectrum for peak at scan number 1000.*

The mass spectrum for the peak at scan number 1000 (RT 16:56mins) is shown in Figure 4.18. The ions present in the spectra are assigned as follows, C_2HF_4Cl *m/z* = 136, C_2HF_3Cl *m/z* = 117, C_2HF_4 *m/z* = 101, CHFCl *m/z* = 67 and CHF_2 *m/z* = 51.

The spectra was compared with previously analysed standards and shown to be CFC 124 (CHF_2CF_2Cl).

The mass spectrum for the final peak in the chromatogram at scan number 1013 (RT 17:09mins) is shown in Figure 4.19. The assignment of ions was as follows, C_2HF_4 *m/z* = 101, $C_2H_2F_3$ *m/z* = 83, C_2HF_2 *m/z* = 63, CHF_2 *m/z* = 51 and CH_2F *m/z* = 33. The library search of this spectrum showed the component to be HFC 134 (CF_3CH_2F). By comparing the mass spectra for HFC 134 and HFC 134a it is clear that even though these two molecules contain the same atoms the positional differences of the atoms within the molecules leads to vastly different mass spectra. This allows isomeric forms of molecules to be clearly distinguishable from one another.

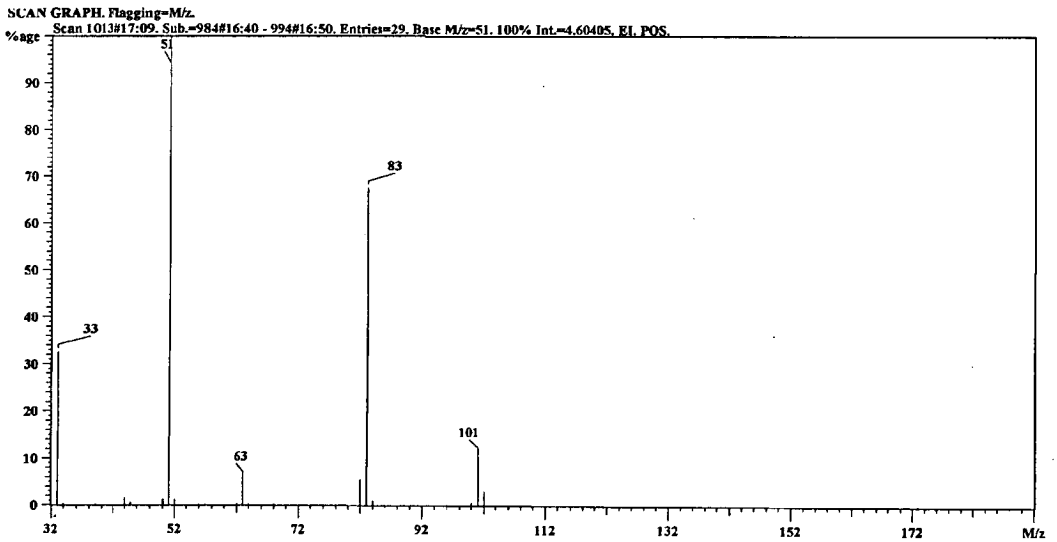


Figure 4.19 *Mass spectrum for peak at scan number 1013.*

The identification of components using mass spectral information allows the peaks to be compared between samples of HFC 134a. A full list of the identified peaks and the comparisons made are listed in Appendix L. As an example, Table 4.2 summarises the identities of the peaks seen in the mass chromatograms obtained from the analyses of four different samples of HFC 134a.

The scan number is used to identify each peak present in the chromatograms and each component is identified in the final column of the table. In some cases the actual identity of a peak is only inferred from the mass spectrum and a definitive identification can not be made. However, where similar mass spectra representing an unknown compound are seen in two separate samples these peaks, although not identified, can be interpreted as the same component. Further, peak identities can be assigned to peaks near the limits of detection (*i.e.* poor quality spectra) if the retention time and any ions present match those of other peaks of better quality spectra (*i.e.* those with a more definitive mass spectrum).

Table 4.2 *Peak identities based on mass spectral and retention time data for the analysed samples of HFC 134a.*

ICI UK	Hoechst	Atochem	Red	**Peak identity**
182	182	182	183	Air
-	-	-	395	CFC 115
-	465	-	-	CFC 12
609	602	-	604	HFC 143a
-	606	-	-	?
-	665	-	-	?
-	680	683	682	CFC 1122
696	685	688	-	HFC 1234
-	696	-	699	HFC 125
-	700	-	-	CFC 22
721	707	712	711	CFC 114
727	713	717	715	CFC 114a
-	717	721	720	?
740	723	728	-	1243
747	728	-	-	?
-	-	733	-	CFC 245
749	729	735	732	HFC 134a
880	884	884	885	HFC 134a + Cl
1000	999	999	1001	CFC 124
1013	1013	1011	1013	HFC 134
-	1091	-	1093	CFC 133

The level of sensitivity obtained for the analyses by mass spectrometry is lower compared with both the FID and ECD traces. The consequence of a reduced level of sensitivity is that the number of overall peaks seen in the mass spectral profiles is less than those seen in both the FID and ECD profiles. The relative sensitivities for

components analysed by FID and mass spectrometry appear similar in that a large component seen in the FID trace will be similarly large in the mass spectral trace. This allows comparison between the FID and MS data to identify the major peaks in the FID trace. The level of sensitivity for the peaks seen in the ECD trace is governed mainly by the number of chlorine atoms contained in the molecule and therefore direct comparisons between ECD traces and MS traces is not possible.

Peaks which are not directly identified from the mass spectral analysis can be tentatively assigned on the basis of retention time in that the larger peaks act as markers in the chromatographic profiles. As the chromatographic order of a number of components is known[11] it is possible to predict a possible identity. For example, the peak seen at scan number 700 in the Hoechst sample appears between peaks corresponding to HFC 125 and CFC 114. From previously analysed standards this peak would be expected to be CFC 22 and by viewing the spectrum, of low quality, the ions present would suggest this component to be CFC 22. From these two observations, neither of which can be conclusive on its own, a more positive identification can be made.

The comparison of the ECD and mass spectral chromatograms cannot be viewed in the same way as the FID/mass spectral work. This is due to the extremely high levels of sensitivity that the ECD shows to molecules which can readily capture electrons and interact with nitrogen ions (Section 1.6.2). Therefore the largest peak in the ECD profile will not always correspond to the most abundant component present in the sample. However, as the mass spectrum of a component containing either chlorine or bromine is very distinctive, due to isotopic ratio patterns, it is possible to closely correlate components identified by mass spectrometry with the large peaks seen in the ECD *i.e.* if a component containing either chlorine or bromine is identified by mass spectrometry a corresponding large peak should be seen in the ECD trace.

4.5 Mixed sample analysis by FID and ECD

The required technique for the analysis and identification of components present in samples of HFC 134a will mainly utilize only GC techniques with FID and/or ECD. The need for the absolute identification of components is not required for classification purposes but it is imperative that the correct peaks are assigned between chromatographic profiles[12]. That is, as long as two peaks are known to represent the same component in different samples it is possible to use these peaks in the classification process without knowing the exact identity of the component.

Identification of components on the basis of retention time or, more accurately, relative retention time, allows peaks to be compared in simple mixture separations. As the number of components in the mixture increases and the retention time values become closer the ability to distinguish between peaks even when they are baseline separated is difficult. By analysing mixtures of samples it is possible to identify the presence or absence of peaks within two samples.

The technique of mixed analysis is demonstrated using two samples of HFC 134a. Each sample of HFC 134a was analysed by GC-FID in triplicate to reduce random error such as ghost peaks and produce a clear and true profile for each sample. The chromatographic profiles for the two samples were compared and areas of ambiguity highlighted. The two samples were then mixed in known ratios (*e.g.* 50:50) using the dilution apparatus. The mixed sample was analysed by GC-FID (n=3) under the same standard conditions and chromatographic profiles obtained.

To interpret the results the following rationale was used:

Presence of by-products	Expected ratio result	
	Ratio with Sample 1	Ratio with Sample 2
In Sample 1 and not Sample 2	approximately 0.5	no value
In Sample 2 and not Sample 1	no value	approximately 0.5
In Sample 1 and Sample 2 where levels are the same	approximately 1	approximately 1
In Sample 1 and Sample 2 where levels in 1>2	slightly greater than 0.5	larger than 0.5
In Sample 1 and Sample 2 where levels in 1>>>2	approximately 0.5	very large ratio

4.5.1 Example of the interpretation of a mixed sample of two HFC 134a

The data representing the early part of the chromatographic profile obtained from the analysis of a mixed sample of HFC 134a are displayed in Table 4.3. The sample consisted of a 50:50 mix (by volume) of two samples of HFC 134a produced by ICI in the USA and in Japan. The mixed sample was analysed using the standard conditions employed for GC-FID.

Table 4.3 *Table of results for mixed sample analysis of USA and Japan samples of HFC 134a by GC-FID (data points taken from beginning of chromatographic trace).*

USA/Japan 50/50 mixed sample			Height response (mV)		Ratio result compared with mixed sample	
Retention time (min)	RRT	Peak height (mV)	USA expected	Japan expected	USA ratio	Japan Ratio
2.64	0.08	127.2	113.0	134.9	1.13	0.94
3.11	0.09	46.98	157.54	0.03	0.30	1381
3.38	0.10	0.12	0.28	-	0.41	-
5.42	0.17	0.03	-	-		-
5.89	0.18	0.03	0.10	-	0.29	-
6.21	0.19	0.41	1.08	-	0.38	-
6.32	0.19	0.54	0.29	0.76	1.83	0.71
8.45	0.26	30.50	16.55	46.88	1.84	0.65
12.65	0.39	5.36	15.94	-	0.34	-
12.78	0.39	0.20	-	-	-	-
14.29	0.44	0.03	0.11	-	0.28	-
15.55	0.47	0.06	0.20	-	0.29	-
17.28	0.53	0.06	0.18	-	0.36	-
20.90	0.64	6.49	13.48	1.91	0.48	3.39
23.08	0.70	0.05	-	0.14	-	0.38
26.21	0.80	0.11	0.22	0.05	0.51	2.47

The first three columns describe the results obtained for the profiling of the mixed sample; column one is retention time, column two is the relative retention time calculated to the leading edge of the peak for HFC 134a and column three is the measured height of the peak in mV on a range setting of 0. Columns four and five show the expected height for each peak seen in the previously analysed samples from the USA and Japan; these results are averaged values of the samples analysed. Columns six and seven are ratio values calculated between the expected heights and the actual heights seen in the mixed sample chromatogram.

The first peak in the mixed sample chromatogram corresponds to the impurity of air in the chromatographic system. This peak would be expected to be seen in both samples of HFC 134a at approximately the same level. The calculated ratios would therefore be expected to be close to 1 for each sample. Looking at the calculated result these values are both close to one and therefore the assignment of the peak as corresponding to the same peak, air, in both samples is correct.

The second peak (3.11) is thought to represent peaks seen in both the USA and Japanese samples. However, the USA sample contains a much higher concentration and should therefore have a ratio of approximately 0.5 whereas the Japanese ratio should be extremely large. The ratios obtained show that the assignment of the peaks was correct.

The third peak (3.38) is thought to represent a component seen only in the USA sample which should therefore give a ratio of approximately 0.5. The result shown for the USA ratio matches what is expected.

The next peak (5.42) is seen in only the mixture and is not expected as neither the USA or Japanese samples were shown to contain this component and it therefore most probably represents a contamination, possibly due to the additional error inherent in mixing the two samples.

The peak at 6.32 has been assigned as being contained in both the Japanese and USA samples. The expected peak height obtained from a 50:50 mixing of the two samples would be approximately 0.53, giving ratio values in the region of 1.9 for the USA sample and 0.7 for the Japanese sample. As can be seen in the table these ratio values are as expected and therefore the assignment of identity between samples is correct.

The other results in the table all follow the same pattern in that the comparison of peaks between samples can be made with more confidence as the linking of peaks is

not made solely on the basis of retention time when no mass spectral information is available. Where samples have vastly different peak heights the comparisons can become difficult as the smaller peaks have no discernible effect on large peaks *i.e.* random variation alters the ratio more than the presence of the smaller peak. However, in these cases it is possible to mix the samples in weighted ratios so that less of the components corresponding to the larger peaks are present *e.g.* in the case of the Hoechst/UK mix performed at 20:80. A full list of the results of the mixed analysis of the samples by GC FID and GC ECD is given in Appendices J and K respectively.

4.6 Data handling

Data acquisition was performed by a PC running the Perkin Elmer 1020 software. The data acquired are stored as files on the hard drive of the PC or on $3^1/_2$ inch floppy disks. The format in which these data files are stored is not directly transferable to the software used to perform either the single variate or multi-variate statistics required in this work. For this reason the peak information was extracted from the chromatographic reports produced by the 1020 software.

The chromatographic report contains the information obtained from the 1020 software's integration of the chromatographic data. The integration parameters are pre-set by the operator to allow automated integration of sample files[13]. The integration parameters include values such as 'peak threshold' which determines the minimum peak area/height required for an integrated peak to be identified as a true peak, and 'baseline noise' which determines the level of background noise present and therefore prevents noise from being integrated as a peak. These parameters can be determined either by calculations based on previous chromatograms *i.e.* measuring levels of baseline drift, or directly by looking at the chromatogram of interest.

Assigning or calculating these parameters is, however, difficult if the chromatogram is particularly complex. With complex separations containing peaks of vastly different sizes and peak shapes which do not always give perfect Gaussian

distribution the setting of the integration parameters becomes difficult[14]. It is often the case that the parameters need to be constantly changed to correctly identify and measure peaks in a chromatogram. The process of automated peak integration becomes time consuming, laborious and prone to error. For these reasons the peaks in this work were manually integrated. Although this is not ideal as the process is still time consuming the data obtained can be treated with a greater degree of confidence and the chromatographic process can be more closely monitored. A restriction in the 1020 software used to acquire the data is that no manual peak integration is possible. Peaks were therefore identified and characterised by their heights. While this is not ideal, the limitations of the available software required the implementation of this method.

The use of peak heights to describe the peaks and hence levels of by-products of synthesis present is subject to a number of constraints. The first and major constraint is on the chromatographic conditions. By altering the temperature gradient components can be eluted at different rates. At higher temperatures components elute more quickly and produce taller narrower peaks. At lower temperatures the peaks are more spread out and smaller in height. Therefore the peak area of a set amount of a component chromatographed under different temperature conditions remains constant but the height will change. For this reason it is imperative that the temperature program used in the chromatographic separation remains constant. It was therefore necessary initially to spend a considerable amount of time on optimising temperature conditions to separate all components in the test samples and allow for any, as yet, undetected potential components. This led to the implementation of a constant temperature profile which had an extended run time (standard conditions detailed in Section 4.2.3).

A second consideration in using peak heights to describe component peaks is the effective linear range. As the amount of a component increases, both the area and height will increase in a linear fashion. However, at extremely high concentrations of components the peak shape can deteriorate due to effective detector range (see

Section 1.8) or from overloading the column. These effects have a greater impact on the measured heights of peaks as opposed to their areas (see Section 4.1.4.3). However, the error within the linearity of the height measurement and area measurement for low levels of components at the LOD stays approximately the same. As the measured levels of the components in this work to characterise the by-product peaks are at low levels it was felt that any possible error in height compared with area would not be significant and certainly reduce errors due to mis-integration of peaks.

The data obtained from the 1020 integrator were manually inputted into an EXCEL spreadsheet. All uni-variate statistical analyses performed on the data was done using EXCEL. The information to describe each sample analysed was taken as a value of retention time correct to three decimal places and a peak height also taken to three decimal places. A value for relative retention time was calculated for each peak in comparison with the leading edge of the peak corresponding to HFC 134a as shown in Equation 4.1[7]. The value was calculated to the leading edge of the peak for HFC 134a as this would remain present throughout the different samples of HFC 134a samples analysed. The retention time of the peak would also reflect any changes in the chromatographic performance of the column, as opposed to relating values to the peak representing air which is present in all the samples but unaffected by column conditions. Data transfer between EXCEL and SPSS was performed by cutting and pasting directly between the two compatible spreadsheets.

$$\text{Relative retention time} = \frac{\text{Retention time of peak}}{\text{Retention time of HFC 134a}} \qquad \textbf{Equation 4.1}$$

The different samples contained a range of different components each at different levels. No one sample contained all the peaks seen in all the samples analysed. For this reason all the samples contained values below the LOD for certain peaks. However, to place a value of zero in the peak height is not necessarily correct as by placing a value of zero in this position is defining this sample as containing absolutely none of the component. The true situation is that the value in any blank position can only be described as being below the LOD of the system *i.e.* if the LOD

is 1pg then the sample could contain 100fg of component which would not be seen; therefore to describe blanks as zeros is incorrect. Therefore, these values below the LOD are given the arbitrary value of half the LOD *i.e.* in the case of the FID detector this value is set at 0.015mV.

References

[1] Mead, R., *The Design of experiments*, Press Syndicate of the University of Cambridge, p7-12, (1992)

[2] Kealey, D. and Fifield, F.W., *Principles and Practice of Analytical Chemistry*, 4th ed., Blackie Academic and Professional, London, p29-30, (1995)

[3] Reineke, F.J. and Bachmann, K., *J. Chromatogr.*, **323**, p323-329, (1985)

[4] Motulsky, H.J., Stannard, P. and Neubig, R., *GraphPad Prism™*, Version 2, GraphPad, San Diego, (1995)

[5] de Zeeuw, J., Zwiep, D. and Marinissen, J.W., *International Laboratory*, **September**, 12J-12P, (1996)

[6] Sturrock, G.A., Simmonds, P.G., Nickless, G. and Zwiep, D., *J. Chromatogr.* **648**, p423-431 (1993)

[7] Fowlis, I.A., *Gas Chromatography*, 2nd ed., John Wiley and Sons, Chichester, England, p166, (1995)

[8] Simmonds, P.G., O'Doherty, Nickless, G., Sturrock, G.A., Swaby, R., Knight, P., Ricketts, J., Woffendin, G. and Smith, R., *Anal Chem*, **67**, p717-723, (1995)

[9] Lee, T.A., *A Beginner's Guide to Mass Spectral Interpretation*, John Wiley and Sons, Chichester, England, Chapter 3, (1998)

[10] Banks, R.E., Smart B.E. and Tatlow J.C., *Organofluorine Chemistry Principals and Commercial Apllications*, 159-175 Plenum Press, New York and London

[11] Handley, A., *Internal Report*, ICI C&P, Runcorn, England, (1994)

[12] Danielson, R. and Malmquist, G., *J. Chromatogr.*, **687**, p71-88, (1994)

[13] *1020 Integrator Users Manual*, Perkin Elmer, England, Chapter 6, (1988)

[14] Dyson, N., *Chromatographic Integration Methods*, Royal Society of Chemistry, Cambridge, p94-101, (1992)

Chapter 5

Data transformation

5.0 Introduction to chapter

This chapter details how the data matrices used to describe the data were formed. These matrices represent the data on which chemometric analyses were performed, as described in Chapter Six. The chapter describes the methods employed to describe the raw data in different formats more amenable to chemometric analysis.

The initial section (5.1) is concerned with the description of the data sets using single variate descriptors which highlights both the usefulness and limitations of such techniques. The second section (5.2) deals with the transformation of the original data sets into different forms. This includes looking at different methods of scaling and normalisation of the data. The more complicated methods of data dimensionality reduction, a method of pre-processing the data for cluster analysis, are dealt with in a separate section (5.3). This is a reflection of the complex nature and the way in which reducing the dimensionality of the data can be viewed as a clustering technique in itself.

5.1 Univariate statistical analysis

The following section describes the application of univariate statistics to the data obtained from the analysis of standards of HFC 134a. Close consideration is paid to the data in order to show that the results obtained from the analyses can be considered to be true descriptions of the samples of HFC 134a. That is, the data produced by the analysis of compounds by GC-FID, ECD or MS can be considered to be of a high standard. This involves mathematically showing that the data obtained are of an accuracy and precision level to allow meaningful conclusions to be drawn from the resulting multi-variate analyses.

5.1.1 Accuracy and precision of measured peaks

Sections 4.1.2 and 4.1.3 showed that the analysis of standard compounds using GC with either FID or ECD gave precise results, in that the variation around a mean value does not alter greatly. Further consideration can be given to the results by investigating the underlying distribution of the data.

The spread of the data for the monitoring of FID and ECD performance can be described in terms of a frequency distribution[1]. This splits the data into regions on the basis of the peak area response. It counts the number of measurements of peak area response which appear in each region and assigns that region a frequency value. A frequency distribution for the analysis of methane standards by FID is shown in Figure 5.1A.

Viewing data in this way is prone to interpretation, since decreasing the width of the frequency domains (Figure 5.1B) flattens the data spread making the results appear less precise. Increasing the width of the frequency domains (Figure 5.1C) sharpens the data spread giving the impression the results are more precise *i.e.* decreases the spread of data.

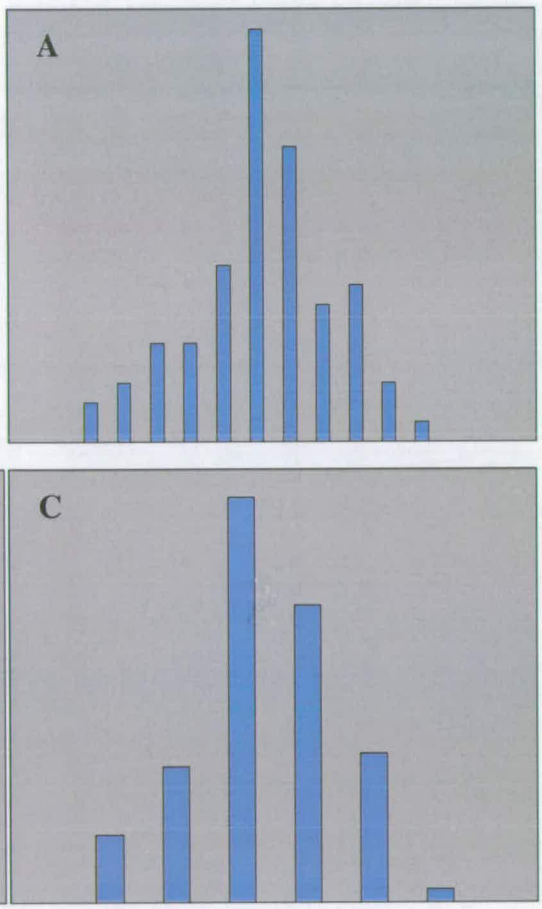


Figure 5.1 *Distribution profiles of results of methane standards analysed by FID with different frequency window sizes over the same range, where area regions C>A>B.*

A more objective measure is to look at the data using normality measurements. The test employed in this work to see if the spread of data deviated from a Gaussain

distribution used a function in the PRISM Statistical package. The function used the Kolmogorov-Smirnov test[2] and calculated a value for P. If the data analysed followed a perfect Guassain distribution the Kolmogorov-Smirnov distance (KS) would equal zero. The larger the KS value the greater the deviation from a Gaussain distribution. PRISM also reported a value for P which answered the question: if the population were really Gaussain, what is the chance that a randomly selected sample of this n number would have a KS value as large as or larger than observed.

To draw a conclusion from the normality test involves considering the P value and the number of observed results. A high P value (> 0.10) can only be considered to indicate a Gaussain distribution if the n number is large. The exact value of large can usually be interpreted as greater than 50-100 observations[3]. Small sample sizes will return large P values whether the data follow a normal distribution or not. The results for the PRISM normality analysis of the FID and ECD stability data are displayed in Table 5.1. The level of confidence applied to the analysis was set at 99% for both the FID and the ECD normality checks.

Table 5.1 *Results of normality testing on the FID and ECD stability data.*

	FID data (peak area counts)	ECD data (peak area counts)
Data		
Number of observations	61	65
Minimum value	282900000	278600000
Maximum value	389400000	406600000
Statistics		
Mean	335400000	359800000
Standard deviation	23700000	26340000
Coefficient of variation	7.07%	7.23%
Normality test		
KS distance	0.118	0.124
P value	P > 0.10	P > 0.10
Passed normality test	Yes	Yes

Both data sets for the FID and ECD stability testing contained >60 samples and could therefore be used to investigate the normality spread of the data. The average value is quoted for each detection method with the percentage relative standard deviation

showing that the results demonstrate a high degree of reproducibility. The calculated P value is greater than 0.1 in both cases and therefore the analysis techniques can be considered as producing data following a normal distribution. This is an important conclusion as it has major consequences on the types of chemometric techniques that can be applied[4].

5.1.2 Accuracy and precision of peak heights for standards of HFC 134a

The chromatographic profiles obtained from the replicate analyses of the standard samples of HFC 134a from the different production sources were investigated using single variate statistics. All the peak heights obtained for each component seen in the different traces had a precision value calculated. This allowed the data to be checked for any abnormal values between each run. The single variate approach shows how each peak differs between the replicate analyses. The precision values should lie within the expected precision limits as determined in Section 4.1.2 and 4.1.3 for both the FID and ECD results (approximately 10% for both detection methods).

The precision values for the different peak heights seen in the different samples of HFC 134a were calculated. As an example, the precision data for peak heights obtained from the replicate analyses (n=5) of the HFC 134a sample from ICI USA are displayed in Table 5.2. The table contains the precision data for the peaks chromatographing before the leading edge of the peak for HFC 134a.

From the table it can be seen that there is a range of precision values (0.8-23.4%). The general trend, however, is that the peaks which show the highest precision, *i.e.* vary little between replicate analyses, are those with large peak heights, that is, peak heights which are considerably greater than the LOD. The peaks which exhibit the greatest variation are those present at the lower concentration levels near the LOD. The smaller peaks have relatively larger precision values because of the effect small variations in the levels of noise or drift can have on the measured height of peaks near the LOD, *e.g.* baseline of peak not always constant between samples. High variation (15%) in peak heights is also seen when peaks overlap especially in the case

of relatively small peaks appearing on the trailing edge of much larger peaks. The reasons for the decrease in precision are again due to measurement factors and not variation in the detectors. A similar effect was seen in the loss of precision for low levels of the standard compounds studied in Section 4.1.4.1-3 when preparing standard lines.

Table 5.2 *Average values for peak heights with precision data for peaks chromatographing before HFC 134a in the replicate analyses of HFC 134a sample from ICI USA.*

Peak Identity	Average RRT	Average height (mV)	Standard deviation of peak heights	Percentage precision of peak heights
Air	0.09	0.539	0.057	10.6
	0.10	1.034	0.054	5.2
	0.14	7.793	0.362	4.6
	0.16	0.163	0.010	6.0
	0.42	4.739	0.382	8.1
	0.43	35.005	1.314	3.8
	0.43	41.086	2.840	6.9
	0.48	0.124	0.029	23.4
	0.58	0.486	0.028	5.8
	0.59	0.077	0.008	10.0
	0.64	0.063	0.006	9.5
	0.70	0.055	0.009	16.1
	0.88	2.356	0.134	5.7
	0.96	0.140	0.001	0.8
HFC 134a	1.00	OS	n/c	n/c

O/S value off scale
n/c value not calculated

The effect of dehydrochlorination by the PLOT alumina column[5] was monitored by looking at the precision values for the different samples of HFC 134a. No dehydrochlorination effect was detected when using the PLOT alumina column. In that case peaks corresponding to low molecular weight hydrohalocarbons were not seen to vary to any greater or lesser extent than other peaks in the chromatographic traces. This is not to say that dehydrochlorination did not take place *i.e.* peak heights may have been reduced to lower levels through a catalytic effect of the column but this would appear to be a consistent effect if it did occur. Therefore any degradation of components because of dehydrochlorination, previously highlighted, does not introduce any extra variation into the analysis.

Components which had peak heights at the LOD led to some chromatographic peaks not being seen in all the replicate analyses. Where peaks were not seen due to their low levels the precision values were calculated using the standard <LOD variable value. This obviously increased the degree of variation for these peaks to high levels; however, the peak values were still used in the characterisation of the samples as all data points are considered legitimate and useable. If peak height values for the standards were ignored due to high levels of variation this would introduce a bias into the modelling system used to characterise and identify samples of HFC 134a. The omission of peaks in the precision calculations on the basis of their presence in only one chromatographic trace (*i.e.* considered as ghost peaks) is discussed in Section 3.6.3.

5.1.3 Calculation of 'typical' chromatographic profile

The calculation of a mean value for each of the peaks seen in any chromatographic trace can be used to form a 'typical' chromatogram. This consists of a list of average peak heights for the different components seen in each sample with averaged values for either retention time or RRT. Tables 5.3 and 5.4 show the averaged results for the replicate analyses (n=5) of each of the standard samples of HFC 134a of known origin by GC-FID. Tables 5.5 and 5.6 show the averaged values for the corresponding analyses performed using GC-ECD. The alignment of peaks between chromatographic traces was made on the basis of the relative retention time data, mixed sample analysis and mass spectral identification of peaks (Appendix L).

These averaged values representing the chromatographs can be displayed as stick diagrams where the x-axis represents time and the y-axis the average height response. A typical stick diagram is shown in Figure 5.2 which corresponds to the values obtained from averaging the data for both peak heights and retention times for the replicate analyses (n=5) of HFC 134a sample from Atochem. The bar representing HFC 134a has been elongated to represent the area that this peak covers in the chromatogram.

Stick diagrams provide a quick reference guide when comparing chromatographic profiles between replicates and samples from different production sources. Stick diagrams also provide an example of the limitations of single variate statistics when applied to complex systems. Comparing one or two averaged chromatographs is possible and in many cases useful but increasing the number of comparisons becomes confusing and difficult. Later sections detail how multi-variate descriptions of the data have overcome these problems and allowed the chromatographic results to be considered in less complex forms.

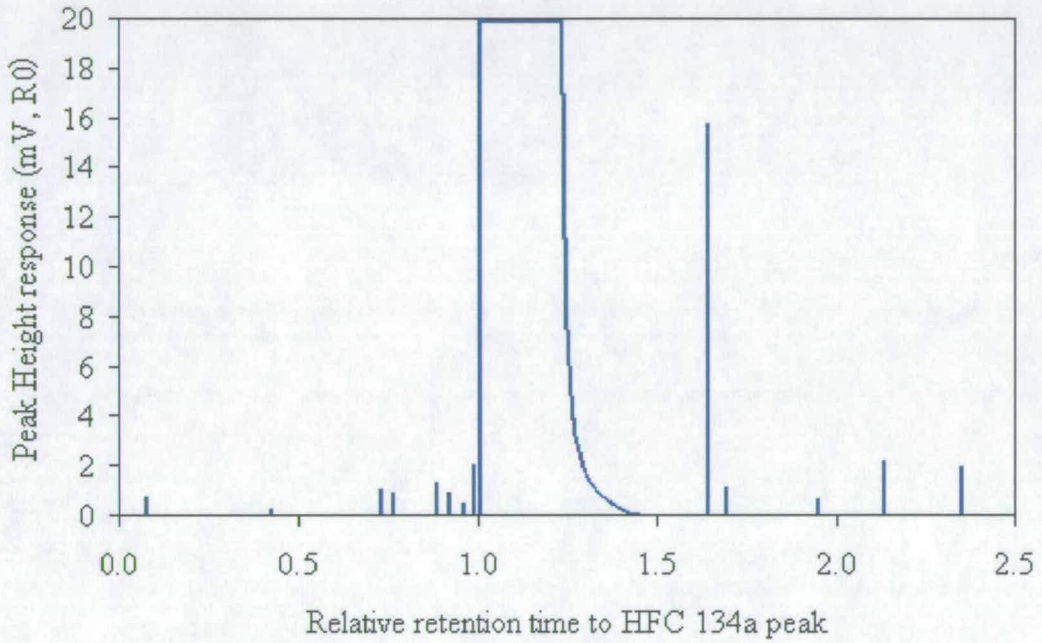


Figure 5.2 *Stick diagram representing average values of peak heights for repeat analyses (n=5) of HFC 134a sample produced by Atochem. Each stick represents a different component (not labelled) present in the sample.*

Table 5.3 *Comparison table for averaged values (n=5) for RRT for known samples of HFC 134a analysed by GC-FID.*

ICI UK	ICI USA	ICI Japan	Atochem	Hoechst	Showa Denka	Ashai Glass	Average RRT
0.086	0.088	0.085	0.089	0.086	0.087	0.086	*0.087*
-	0.101	-	-	0.108	-	-	*0.105*
-	-	-	-	0.111	-	-	*0.111*
-	-	-	-	0.115	0.116	-	*0.116*
-	0.139	-	-	0.137	-	-	*0.138*
-	0.160	-	-	-	-	-	*0.160*
-	-	-	-	0.209	-	-	*0.209*
-	-	-	-	0.240	-	-	*0.240*
-	-	-	-	0.256	-	-	*0.256*
-	-	-	-	0.282	0.292	-	*0.287*
-	0.424	-	-	-	-	0.424	*0.424*
-	0.427	-	-	-	-	-	*0.427*
0.429	0.432	-	0.440	0.428	0.436	-	*0.433*
-	0.477	-	-	0.467	-	-	*0.472*
-	0.578	-	-	-	-	-	*0.578*
-	0.594	-	-	-	-	-	*0.594*
-	0.637	-	-	-	-	-	*0.637*
-	0.696	-	-	0.691	-	-	*0.694*
0.730	-	-	0.740	0.734	0.738	-	*0.736*
0.766	-	0.766	0.775	0.768	0.773	-	*0.770*
0.859	0.861	-	-	0.859	-	-	*0.860*
0.885	-	0.891	0.894	0.880	-	-	*0.888*
-	-	-	-	0.887	-	-	*0.887*
0.917	-	-	0.928	0.923	-	-	*0.923*
-	0.964	-	0.958	0.959	-	-	*0.960*
0.962	-	0.960	0.963	0.964	0.964	-	*0.963*
-	-	-	-	0.991	-	-	*0.991*
0.995	-	-	0.996	0.995	-	-	*0.995*
1.000	1.000	1.000	1.000	1.000	1.000	1.000	*1.000*
1.662	1.637	1.621	1.664	1.659	1.646	1.623	*1.645*
1.709	1.681	-	1.713	1.761	1.669	1.663	*1.699*
-	2.021	1.810	-	1.899	-	-	*1.910*
1.918	-	-	1.924	2.017	-	-	*1.953*
2.157	2.165	2.102	2.144	2.122	2.138	2.127	*2.136*
-	-	2.124	-	-	2.172	2.161	*2.152*
-	2.221	2.160	-	-	-	-	*2.191*
2.351	2.551	2.271	2.374	2.305	2.322	2.309	*2.355*
2.353	-	2.327	-	-	-	2.331	*2.337*

Note that because a number of component peaks elute at very similar times and chromatographic retention times vary between runs some of the RRTs appear reversed *e.g.* 0.888 and 0.887 because of the effect of averaging across the samples.

Table 5.4 *Comparison table for averaged values (n=5) for peak heights for known samples of HFC 134a analysed by GC-FID.*

ICI UK	ICI USA	ICI Japan	Atochem	Hoechst	Showa Denka	Ashai Glass	Average RRT
0.646	0.539	0.589	0.566	0.489	0.477	0.561	*0.087*
-	1.034	-	-	0.132	-	-	*0.105*
-	-	-	-	2.952	-	-	*0.111*
-	-	-	-	0.119	0.193	-	*0.116*
-	7.793	-	-	0.078	-	-	*0.138*
-	0.163	-	-	-	-		*0.160*
-	-	-	-	0.149	-	-	*0.209*
-	-	-	-	0.197	-	-	*0.240*
-	-	-	-	0.515	-	-	*0.256*
-	-	-	-	8.303	0.057	-	*0.287*
-	4.739	-	-	-	-	0.667	*0.424*
-	35.005	-	-	-	-	-	*0.427*
3.328	41.086	-	0.095	24.310	0.587	-	*0.433*
-	0.124	-	-	23.279	-	-	*0.472*
-	0.486	-	-	-	-	-	*0.578*
-	0.077	-	-	-	-	-	*0.594*
-	0.063	-	-	-	-	-	*0.637*
-	0.055	-	-	1.730	-	-	*0.694*
0.488	-	-	0.881	0.343	0.051	-	*0.736*
0.081	-	0.241	0.736	4.775	0.045	-	*0.770*
0.063	2.356	-	-	0.760	-	-	*0.860*
1.799	-	0.206	1.093	7.028	-	-	*0.888*
-	-	-	-	7.900	-	-	*0.887*
3.320	-	-	0.743	4.199	-	-	*0.923*
0.000	0.140	-	0.113	0.206	-	-	*0.960*
0.939	-	0.271	0.321	0.409	0.046	-	*0.963*
-	-	-	-	0.118	-	-	*0.991*
0.531	-	-	1.832	0.661	-	-	*0.995*
1.000	1.000	1.000	1.000	1.000	1.000	1.000	*1.000*
7.047	2.837	0.706	15.525	8.395	0.448	1.374	*1.645*
2.301	0.402	-	0.938	1.254	0.165	0.168	*1.699*
-	3.302	3.444	-	19.650	-	-	*1.910*
0.360	-	-	0.485	5.145	-	-	*1.953*
11.160	4.044	4.420	2.015	2.386	2.117	2.593	*2.136*
-	-	0.132	-	-	0.092	0.113	*2.152*
-	1.613	0.259	-	-	-	-	*2.191*
3.406	1.929	1.974	1.773	1.724	1.732	1.908	*2.355*
1.745	-	1.700	-	-	-	2.870	*2.337*

Table 5.5 *Comparison table for averaged values (n=5) for RRT for known samples of HFC 134a analysed by GC-ECD.*

ICI UK	ICI USA	ICI Japan	Atochem	Hoechst	Showa Denka	Ashai Glass	*Average RRT*
0.083	0.080	0.077	0.081	0.081	0.084	0.077	*0.080*
-	0.095	0.091	-	-	-	-	*0.093*
-	0.103	-	-	0.105	-	-	*0.104*
-	0.177	-	-	-	-	-	*0.177*
0.199	0.195	-	-	0.199	0.199	-	*0.198*
0.203	0.198	0.192	0.203	0.202	0.206	0.190	*0.199*
0.267	-	-	-	0.268	0.271	-	*0.269*
0.272	0.267	0.260	0.275	0.272	0.276	0.257	*0.269*
-	0.399	-	-	-	-	-	*0.399*
-	-	-	-	0.455	-	-	*0.455*
-	0.495	-	-	0.573	-	-	*0.534*
-	0.548	-	-	-	-	-	*0.548*
0.672	0.663	0.655	-	0.678	0.677	-	*0.669*
-	-	0.729	0.681	-	-	-	*0.705*
-	-	-	0.757	0.754	-	0.650	*0.721*
-	0.824	0.829	-	-	-	-	*0.826*
0.846	0.835	-	0.854	-	-	-	*0.845*
0.869	0.867	-	0.882	0.877	-	-	*0.874*
0.884	0.876	0.863	0.892	0.890	0.880	-	*0.881*
-	-	0.873	-	-	-	-	*0.873*
0.907	0.905	0.903	0.918	0.914	0.919	0.868	*0.905*
0.998	-	0.936	0.932	0.996	-	0.913	*0.955*
1.000	1.000	1.000	1.000	1.000	1.000	1.000	*1.000*
1.219	1.211	1.165	1.262	1.216	1.263	1.130	*1.209*
1.560	1.520	-	-	-	-	1.495	*1.525*
-	-	-	-	-	1.643	1.505	*1.574*
1.601	-	-	1.664	1.628	1.668	-	*1.640*
-	1.892	1.820	-	1.853	1.921	1.728	*1.843*
1.840	-	1.904	1.893	1.867	1.980	-	*1.897*
-	1.962	1.913	-	-	2.073	1.785	*1.933*
1.904	-	-	1.951	2.007	2.084	-	*1.987*
-	-	-	2.035	2.023	2.087	1.868	*2.003*
-	-	-	2.062	-	2.110	1.877	*2.016*
1.957	-	-	-	-	-	1.876	*1.916*
2.044	2.008	1.929	2.082	2.041	2.106	1.897	*2.015*
2.071	2.034	1.954	2.104	2.066	2.133	1.919	*2.040*
-	2.046	1.965	-	2.078	2.145	1.932	*2.033*
2.082	-	1.981	2.116	-	2.157	1.943	*2.056*
-	-	-	2.127	-	-	-	*2.127*
2.103	2.067	1.985	2.138	2.099	2.166	1.952	*2.073*
2.193	2.153	2.070	2.224	2.188	2.257	2.035	*2.160*
-	-	-	2.232	-	2.263	-	*2.248*
-	-	-	-	-	2.325	-	*2.325*

Table 5.6 *Comparison table for averaged values (n=5) for peak heights for known samples of HFC 134a analysed by GC-ECD.*

ICI UK	ICI USA	ICI Japan	Atochem	Hoechst	Showa Denka	Ashai Glass	*Average RRT*
115.6	120.2	114.1	72.12	85.16	92.92	138.6	*0.080*
-	156.8	0.03	-	-	-	-	*0.093*
-	0.28	-	-	2.06	-	-	*0.104*
-	0.10	-	-	-	-	-	*0.177*
0.14	1.07	-	-	82.94	0.03	-	*0.198*
0.97	0.29	0.66	0.21	3.53	0.19	0.16	*0.199*
0.18	-	-	-	1.11	0.04	-	*0.269*
3.12	16.51	40.00	1.31	31808	18.25	1.22	*0.269*
-	15.90	-	-	-	-	-	*0.399*
-	-	-	-	3.72	-	-	*0.455*
-	0.20	-	-	0.08	-	-	*0.534*
-	0.17	-	-	-	-	-	*0.548*
79.10	13.40	1.63	-	518.1	0.97	-	*0.669*
-	-	0.12	0.93	-	-	-	*0.705*
-	-	-	0.57	4.63	-	0.13	*0.721*
-	0.22	0.04	-	-	-	-	*0.826*
0.47	22.20	-	0.20	-	-	-	*0.845*
493.3	1.60	-	278.4	882.4	-	-	*0.874*
1058	1.89	0.35	16.94	35.00	0.72	-	*0.881*
-	-	0.07	-	-	-	-	*0.873*
3605	10.89	5.20	882.2	4874	14.78	0.30	*0.905*
1.92	-	0.09	0.84	2066	-	2.95	*0.955*
1.55	1.55	1.40	1.33	1.52	1.48	1.54	*1.000*
10.33	21.50	3.96	2.94	977275	11.47	5.45	*1.209*
1.40	154.4	-	-	-	-	0.34	*1.525*
-	-	-	-	-	0.46	1.05	*1.574*
22.10	-	-	7.80	60.00	1.55	-	*1.640*
-	1.08	0.90	-	158.4	44.55	0.93	*1.843*
9.33	-	0.07	6.31	25.75	1.19	-	*1.897*
-	0.26	0.24	-	-	0.24	0.87	*1.933*
1.48	-	-	0.94	4.83	0.28	-	*1.987*
-	-	-	0.09	21.52	0.13	0.07	*2.003*
-	-	-	0.26	-	0.06	0.19	*2.016*
3.68	-	-	-	-	-	0.07	*1.916*
5.47	3.17	1.99	1.84	2.03	3.78	1.81	*2.015*
445.1	400.7	340.1	356.4	341.6	371.2	300.8	*2.040*
-	4.39	3.30	-	38.31	24.27	4.11	*2.033*
3.41	-	0.25	13.06	-	2.38	0.21	*2.056*
-	-	-	1.07	-	-	-	*2.127*
189.6	181.0	132.2	146.0	172.9	198.1	185.5	*2.073*
12.08	10.08	7.86	30.09	111.3	70.68	7.87	*2.160*
-	-	-	1.49	-	1.34	-	*2.248*
-	-	-	-	-	1.04	-	*2.325*

5.2 Data transformation

The following section describes the different ways the raw data were processed before they were subjected to the various mathematical classifications used in this work. A definition of each data transformation is given followed by a short note on the specific pros and cons of the mathematical technique as applied to the chromatographic data investigated in this work.

5.2.1 No pre-processing of the data

The data for both the FID and ECD analysis of the different samples of HFC 134a can have no pre-processing performed on them. That is, the values of the measured peak heights can be used directly in the chemometric classification of the samples. By not performing any pre-processing on the data the classification obtained will directly reflect the results obtained.

Considering the objectives of this thesis this at first glance would appear to be both the simplest and most logical way to treat the data and in many cases of multi-variate classification this would be true. However, using only the raw data to describe the samples takes no account of the differences in relative peak heights between the different by-products present. That is, is it a correct assumption that just because a by-product is present at high levels it is the most important peak in the chromatogram? The largest peak may represent the most important component when considering the purity of a sample but it does not follow that for classification purposes this peak is crucial to define a sample. It is likely that the larger peaks will obscure the subtle differences present between smaller peaks because of scaling factors and not actual differences or similarities in the samples. However, transforming the data does lead to the question that any classification based on transformed data may be subject to some bias. That is, groups are formed because of the way the data have been treated before examination by chemometric techniques. It is therefore necessary to view with care each result obtained using the different chemometric techniques in order to avoid a biased characterisation method[6].

5.2.2 Scaling of the data

Pre-processing of the data by some form of scaling is performed in order to remove effects on the classification because of the magnitude of peak heights. By scaling the values for peaks heights it is possible to reduce the effect of the absolute size of peaks. Reducing the effect of absolute size means that the classification will be based more on the inter-relations between peak heights and not be biased by a few larger peaks present. However, in scaling the peaks a certain amount of the variability of the data was removed. It may, after all, be very important to the classification of the samples of HFC 134a which peaks were present at the highest levels.

In the case of the data sets for the FID and the ECD it is apparent that scaling the data will have a larger effect on the ECD data. This is because the range of absolute peak heights in the ECD chromatographic profiles is much greater than in the FID profiles.

The data can be scaled in a number of different ways[6]. The method that was employed in this work was to scale the value for each peak height, in a series of chromatographic traces, to range between 0 and 1. To obtain these values all the data to be used from each repeat analysis of the samples of HFC 134a were considered together in one matrix. The value representing the highest peak height for each component detected was located. All peak heights for each by-product present were then divided by their corresponding highest value. This returns a value with a possible minimum of 0 and a maximum of 1 (the largest peak divided by itself).

Table 5.7 shows an example of a series of peaks from repeat analyses of three samples of HFC 134a produced by ICI in UK, USA and Japan that were subjected to the scaling process described. The values for the peak heights for three by-products seen in the three samples of HFC 134a are shown in the columns labelled Peak A, Peak B and Peak C (the actual identity of the peaks is unimportant). The number printed in bold represents the largest peak height seen in any of the nine chromatograms viewed for each by-product. The values calculated for the

transformed data are each of the original peak heights divided by the peak height value in bold. Transforming the data in this manner does not affect the inter-peak heights (Peak B for Japan is still relatively small in comparison with Peak B for USA after the transformation). However, the intra-peak heights have changed dramatically. The relative heights of Peak A and Peak C before and after the transformation have changed so that any measurements made to classify similarities/differences between the sample will not be affected by the differences in absolute magnitudes.

Table 5.7 *Peak height data for three peaks seen in ECD chromatographic profile for three samples of HFC 134a before and after scaling transformation.*

Sample analysed		Raw data peak heights (mV)			Transformed data (scaled 0-1)		
		Peak A	Peak B	Peak C	Peak A	Peak B	Peak C
UK	Repeat 1	116.60	0.015	0.836	0.802	0.0001	0.798
	Repeat 2	116.00	0.015	0.968	0.798	0.0001	0.924
	Repeat 3	105.10	0.015	**1.048**	0.723	0.0001	1.000
USA	Repeat 1	116.00	**164.00**	0.296	0.798	1.0000	0.282
	Repeat 2	116.20	158.50	0.324	0.799	0.9665	0.309
	Repeat 3	135.90	136.60	0.304	0.935	0.8329	0.290
Japan	Repeat 1	132.70	0.032	0.762	0.913	0.0002	0.727
	Repeat 2	128.40	0.042	0.798	0.883	0.0003	0.761
	Repeat 3	**145.40**	0.023	0.741	1.000	0.0001	0.707

5.3 Reduction of the dimensionality of the data

The following section details the steps taken to reduce the dimensionality of the data obtained from the chromatographic profiling of the known samples of HFC 134a. That is, form linear combinations of the original variables (peak heights) so that these new variables describe the majority of the data's variance in a greatly reduced number of variables (three or less). The section details the methods used and results obtained by employing these data reduction methods in order to reduce the number of required descriptive variables[7,8]. The chromatographic traces for the FID and ECD data are treated separately and also as a combined data set in order to try to fully characterise the different samples of HFC 134a examined.

5.3.1 Extraction of principal components for FID chromatographic data

As detailed in Section 4.6 all multi-variate data handling and investigation was performed using SPSS (version 9.0 for Windows). The data were transferred from an EXCEL spreadsheet directly into the SPSS spreadsheet. The SPSS spreadsheet was set up so that the different samples analysed were represented as the rows and the different peaks identified as the columns. Therefore in all the notation of SPSS the different samples are termed as cases and the peaks are defined as the variables. The data set used for the principal component (PC) analysis of the FID data for the known samples of HFC 134a consisted of thirty five cases (samples) from seven known sources which were described by thirty eight variables (peaks) where each value represented a peak height. The values for peak 29 (HFC 134a) were not used in the calculation as these were off scale.

5.3.1.1 PC extraction from the variance/co-variance matrix

The raw data was subjected to principal component analysis (PCA) by calculating the eigenvalues and eigenvectors from the variance/co-variance matrix[9]. This method initially calculates the variance/co-variance matrix of each variable with every other variable which is used to describe the data. The variance/co-variance matrix is a square matrix which is symmetrical about the diagonal i.e. the variance between Variable 5 and Variable 8 is the same as the variance between Variable 8 and Variable 5. The matrix diagonal represents the value for the variance within any one variable. It is important to note that the variance/co-variance matrix makes no adjustments for the relative magnitudes of the different values used to describe the data. That is, the system is under more influence from the larger peaks in the FID chromatographic trace.

The number of new variables (components) which can be formed to describe all the variation seen within the data must equal the original number of variables (thirty seven). However, whereas the original variables describe variance relating to a specific peak, the principal component variables describe variance relating to linear combinations of these peaks[10]. This means that the first principal component (the

eigenvector with the largest eigenvalue) describes more of the original variation than any of the original variables. The amount of variation described by each of the principal components can be measured by comparing the eigenvalues. As the eigenvalues are a measure of the magnitude of the eigenvectors which are the axes onto which the variance/co-variance calculations were transformed, they can be used to calculate the variance contained within any one principal component[4].

Table 5.8 lists the first ten principal components and their corresponding amounts of total variance related to the data set analysed. The calculation of the percentage variance is performed by dividing each individual eigenvalue for a principal component by the sum of all the eigenvalues. It is immediately apparent that after the first four principal components the amount of variance described thereafter by each successive component is small and probably accounts for random background noise and not variations between samples[11]. The values for each components variance can be plotted onto a graph (a scree plot) to give a visual representation of the degree of variance described by each component. Figure 5.3 shows the scree plot for the first twenty components extracted from the FID data. The figure shows that the relative amount of variation (eigenvalue) is small after component four.

Table 5.8 *Variance described by first ten components extracted from the variance/co-variance matrix of FID data for known samples of HFC 134a*

Principal component	Eigenvalue	Percentage variance	Cumulative variance
1	381.8	64.10	64.10
2	169.2	28.42	92.52
3	26.06	4.38	96.89
4	14.56	2.44	99.34
5	1.409	0.24	99.58
6	1.147	0.19	99.77
7	0.657	0.11	99.88
8	0.372	0.06	99.94
9	0.116	0.02	99.96
10	0.075	0.01	99.97

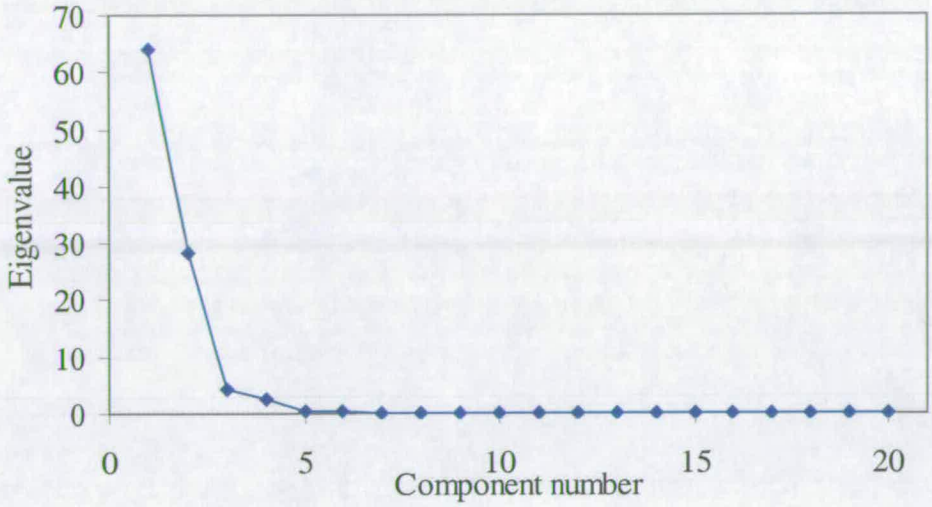


Figure 5.3 *Scree plot for the components extracted from the FID chromatographic data for known sample of HFC 134a using the variance/co-variance matrix.*

The next stage in the PCA is to transpose the original data onto the eigenvectors, that is calculate component scores for each sample for each component[12]. These scores will represent the new values by which each sample is described. The scores are calculated by crossing the original data set with the component matrix. This produces a score for each sample in each of the thirty seven components. The component scores for the first ten components have been calculated and are shown in Appendix M. As described above, the majority of the variance within the data can now be described using a greatly reduced number of variables.

The component scores for each sample can be plotted onto a graph in the same way as any value for a variable. Figure 5.4 shows the values for the first principal component (PC 1) plotted against an arbitrary value (values set the same for each sample from the same source). The figure clearly shows that the samples from the USA and Hoechst have been separated from the other samples and each other. The values for the other samples, although they appear precise, do overlap with one another. This means that these samples cannot be distinguished from each other on the basis of PC 1 and hence are classified as one group.

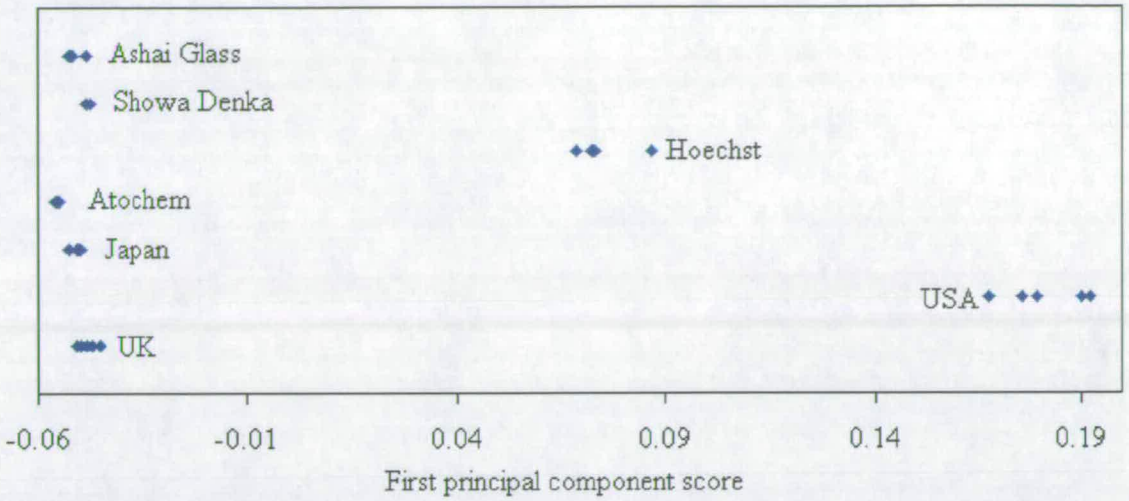


Figure 5.4 *First principal component plotted against an arbitrary value for FID data of known samples of HFC 134a.*

The second principal component can be similarly plotted (Figure 5.5) against an arbitrary value to obtain a separation pattern based on the variation contained within this new variable. The variation described by PC 2 again shows the Hoechst and USA samples are the most different when compared with the other samples. However, it is now possible to separate the Atochem and UK samples from the Japan, Showa Denka and Ashai Glass samples.

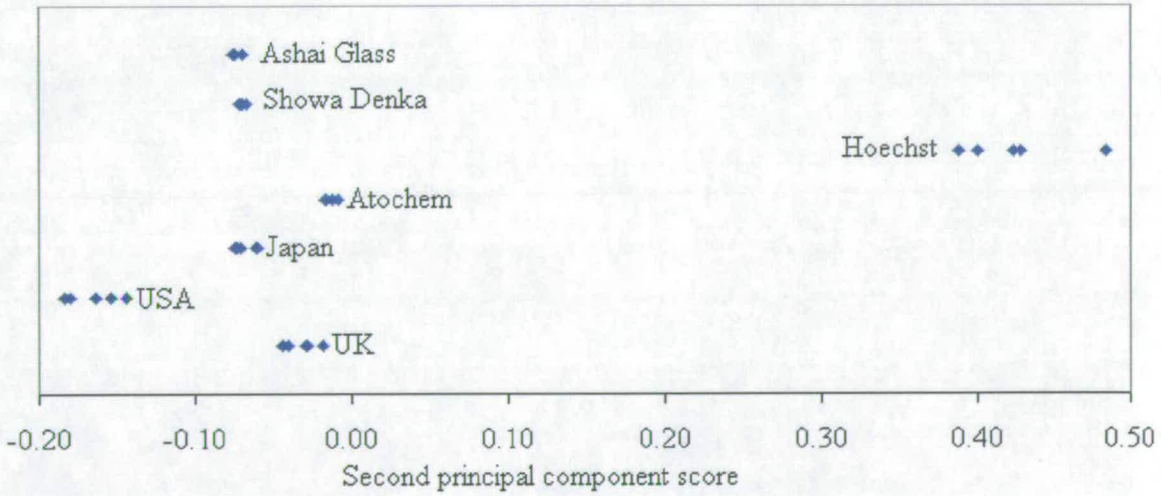


Figure 5.5 *Second principal component plotted against an arbitrary value for FID data of known samples of HFC 134a.*

Plotting the third principal component against an arbitrary value gives the separation shown in Figure 5.6. In this figure it is clear that both the Atochem and UK samples are distinct from the other samples. It should be noted from the eigenvalues in Table

5.8 that PC 3 describes only 4.38% of the total variance when extracting the PC from the variance/co-variance matrix. However, this variance is enough to allow a number of different samples to be distinguished from one another (*e.g.* Japan from Hoechst). As the amount of variance described by PC 3 is relatively small the level of background noise and random variability is beginning to have more effect on the results. This can be seen as the points representing a number of the samples are beginning to spread out, particularly noticeable in the case of the Hoechst and USA samples.

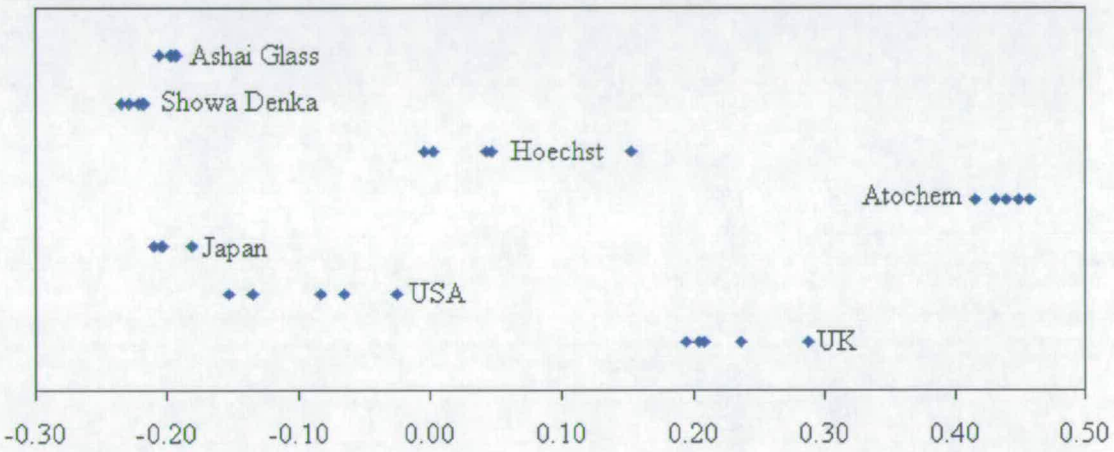


Figure 5.6 *Third principal component plotted against an arbitrary value for FID data of known samples of HFC 134a.*

Looking at the separation patterns so far it is clear that by considering the first three principal components alone it is possible to distinguish clearly between a number of the different samples. However the samples from Japan, Showa Denka and Ashai Glass have not yet been clearly separated from one another. It is therefore necessary to continue to look at further components to aid the separation of these samples. Considering more than three components at any one time becomes difficult as the visual representations are limited. However, the multi-variate classification techniques (Section 6.1) will allow the consideration of many variables mathematically. It should be borne in mind, however, that this is a method of reducing the dimensionality of the data and increasing the number of components considered above a certain number becomes a contradictory exercise.

The obvious extension of these comparisons is to view each component not as an individual variable but in conjunction with the other components. By viewing the PCs together more of the original variation can be viewed in one diagram. In the example given in Figure 5.7a PC 1 is plotted against PC 2. This one graph describes approximately 93% of the original variance in the chromatographic data obtained from the analysis of the samples using FID.

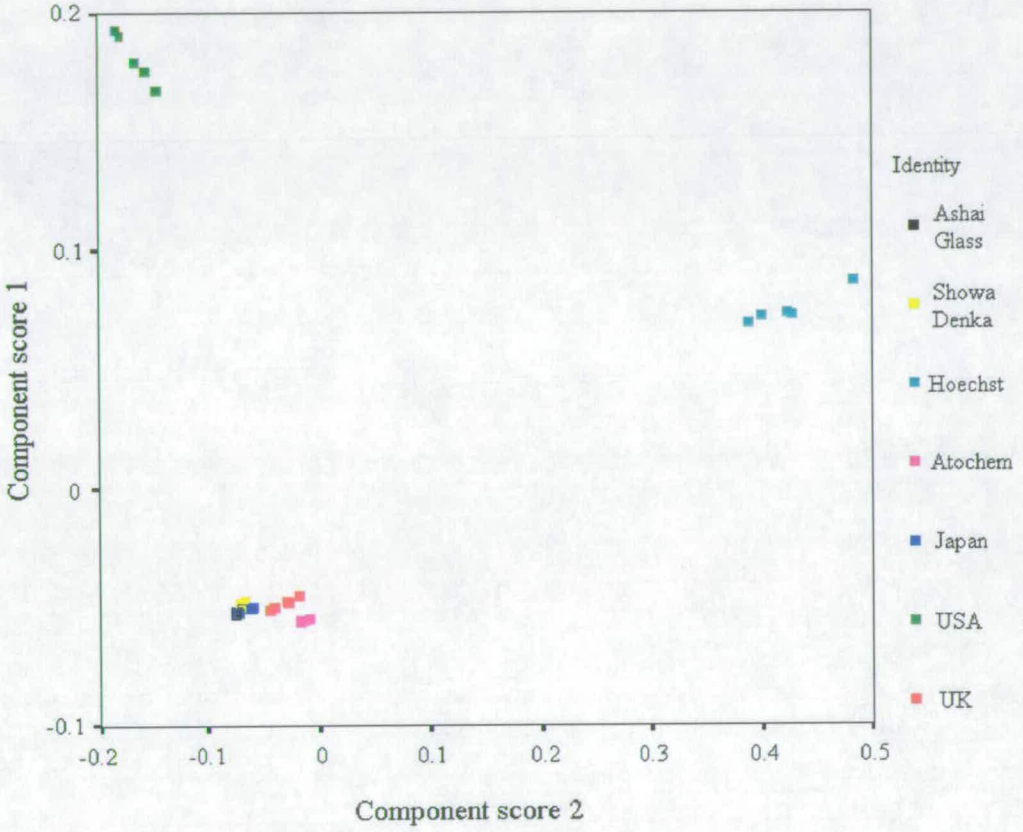


Figure 5.7a *Principal component 1 and 2 plotted against each other for FID data of known samples of HFC 134a.*

The Hoechst and USA samples are clearly distinguished from the rest of the samples. The figure appears to show that the 'rest' of the samples are clustered into one group. However, by enlarging the scale in the region these samples are clustered together (see Figure 5.7b) it becomes apparent that they are in fact split into areas representing the different production origins. By viewing the data using the first two components all the samples of HFC 134a from the different production origins can be separated to a certain degree. The samples representing Japan, Showa Denka and Ashai Glass still overlap.

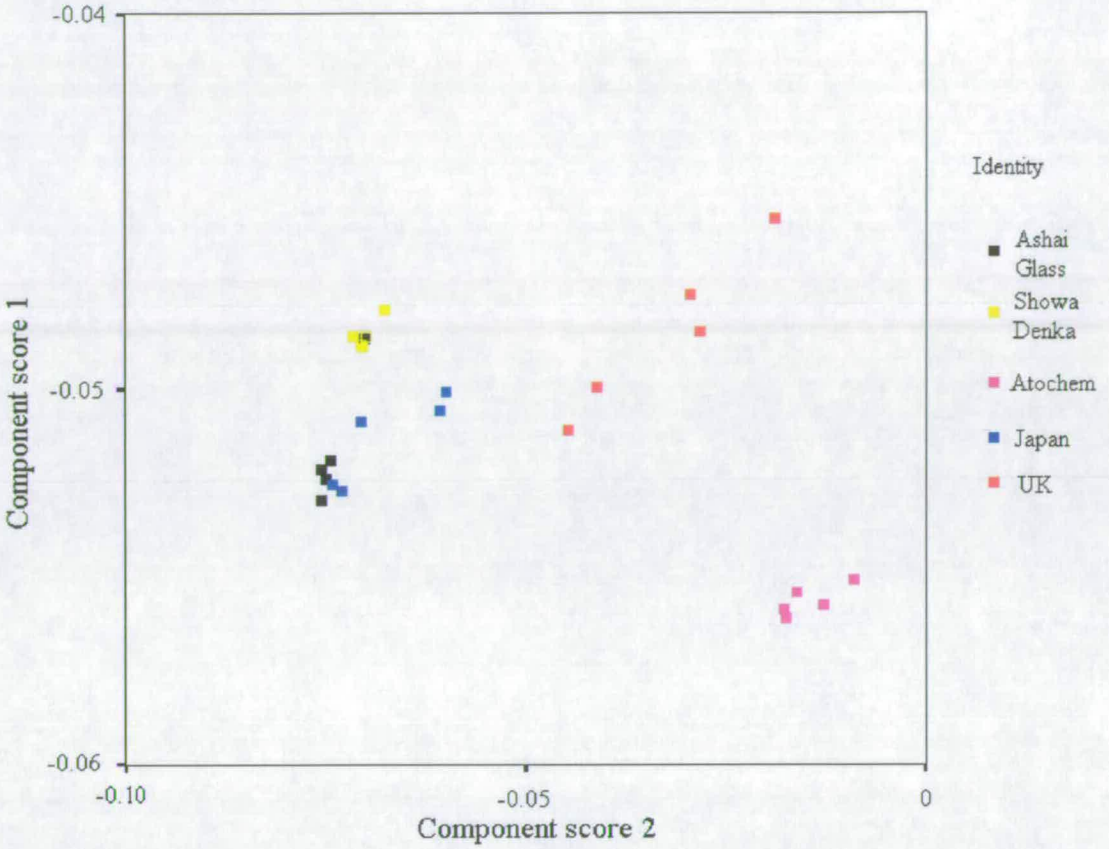


Figure 5.7b *Enlarged scale in the 'rest' region for principal component 1 and 2 plotted against each other for FID data of known samples of HFC 134a.*

The first three components can be plotted onto three dimensional axes where each axis represents a new variable. This graph (Figure 5.8) describes approximately 97% of the total variation of the original data and shows a greater level of separation of the samples. The samples representing UK, USA, Atochem and Hoechst are all clearly separated into distinct groups. However, the sample groups for Japan, Showa Denka and Ashai Glass are still positioned closely together and in a number of cases the points overlap within the groups.

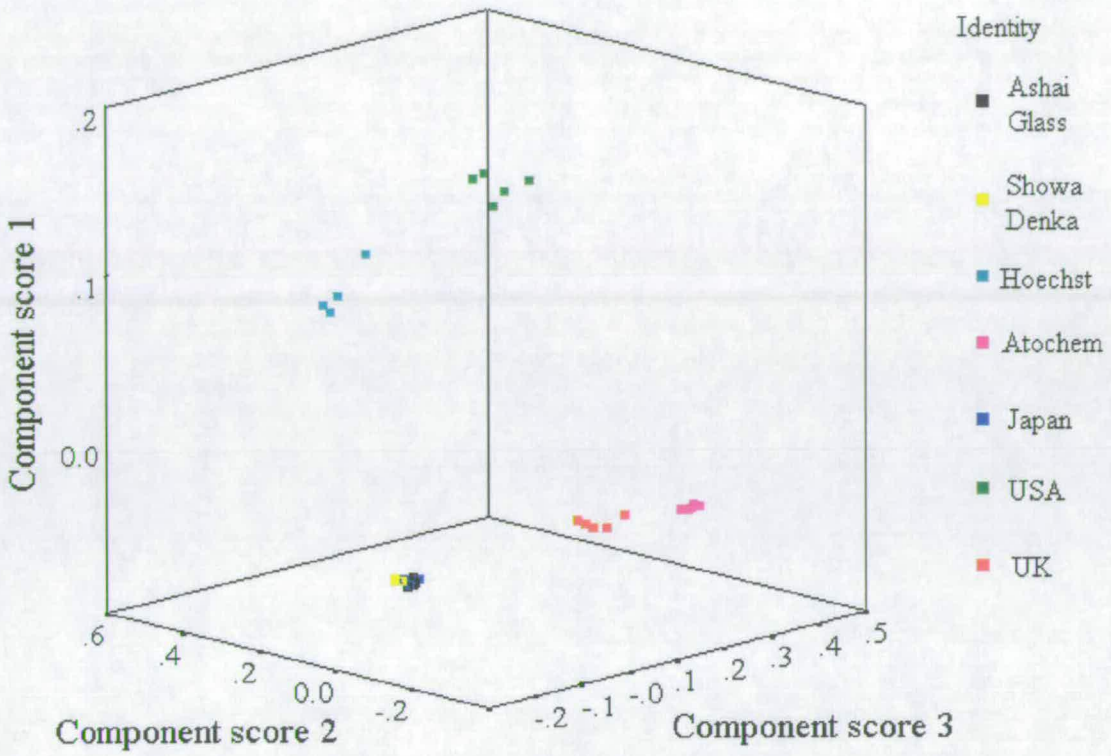


Figure 5.8 *Three dimensional plot for principal component 1, 2 and 3 plotted against each other for FID data of known samples of HFC 134a*

The principal components that have been used to describe the data can be further investigated to determine which linear combination of variables has been used to make up each component; that is, which of the original variables (peaks) are important in the formation of the components. The peaks which have the most effect on the different components are those peaks which are most important in the characterisation of the different samples. As the peaks represent the presence or absence of different by-products of synthesis the information is invaluable when trying to assign some reasoning to why samples from different production locations are similar or dissimilar.

The component matrix contains the values which relate to the influence each variable has upon the different components. The actual raw values for the component matrix used to form the principal components extracted from the variance/co-variance matrix can be re-scaled. The re-scaled values range from 1 (total positive correlation) through 0 (no correlation) to –1 (total negative correlation). Any peak

with a value approaching either 1 or -1 in a component can be considered as having a major influence on that component. Viewing the results graphically (Figure 5.9) allows the quick interpretation of which by-products are exerting a major effect on the principal components and therefore which peaks are the most important in the separation of the samples.

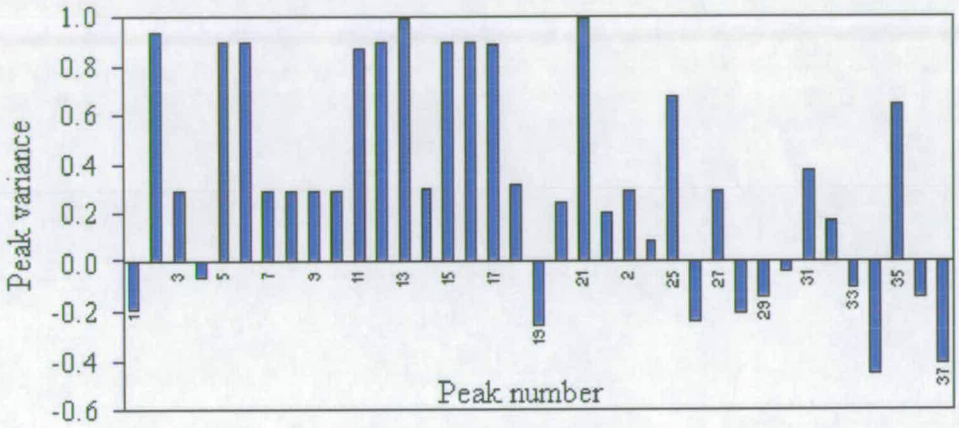


Figure 5.9 *Levels of effect each peak in the FID chromatographic trace has on component 1.*

From the figure it is apparent that Peaks 2, 5, 6, 11, 12, 13, 15, 16, 17 and 21 are all important in the description of PC 1. As the values for these peaks are all positive they have a positive correlation with PC1. There are no peaks with major negative correlations. In the case of Peak 31, as the value in the component matrix is near to zero, it has no effect on PC 1 and therefore does not affect the component score. This means that this peak adds no information to any of the discriminations seen between samples when viewing PC 1 in isolation *i.e.* removing this peak from the chromatographic profile does not affect the separation seen in Figure 5.4.

The peaks which have the major influence on the separation of the samples on the basis of PC 1 can also be considered as those which contain the majority of the variation within the data. This is because these peaks contribute to the greatest extent on the PC which contains the most variation (64%). However, as detailed earlier in this section the extraction of principal components from the variance/covariance matrix makes no allowances for the differences in absolute magnitudes of peaks. That is, the peaks with the relatively largest heights will have more influence on the

formation of the components than those peaks with the relatively smaller heights. For this reason the majority of the variation both within and between samples is because of the absolute peak heights. This reduces the importance of the relatively smaller peaks which in turn reduces the impact of the more subtle differences between the samples. This weighting of importance can in some multi-variate analyses be acceptable, as it is not unreasonable to consider the largest peaks as the most important in the chromatograph. However, the chromatographs produced by the FID have a range of heights and the presence or absence of peaks, no matter the relative height, is considered important in this work. Therefore the use of the variance/covariance matrix to extract the principal components is of limited application in the work presented.

5.3.1.2 PC extraction from correlation coefficients matrix

A second method of calculating the principal components is to extract the eigenvectors from the matrix of correlation coefficients (see Section 2.1.5). The variance/covariance matrix is calculated, as with the first technique, producing a square matrix symmetrical about its diagonal. However, each variance/covariance calculation is then divided by the product of their standard deviations. As the value for the covariance can equal but never exceed the product of the standard deviations the values will range from 1 to -1. Thus the use of the correlation coefficients allows the estimation of the degree of interrelation between variables free from the effects of absolute peak heights. Values approaching 1 indicate a high positive interdependence between two variables, whereas a negative value indicates that one variable decreases as the other increases. Values approaching 1 can be viewed as linearly independent of each other.

The raw data for the chromatographic profiling of the known samples of HFC 134a by FID were subjected to a principal component analysis using SPSS. The matrix for the correlation coefficients contains values of covariance which range from –1 to 1. The matrix diagonal, which represents the correlation between a variable and itself, has a value of 1 by definition.

The eigenvalues for the first ten principal components are shown in Table 5.9 along with percentage variance described by each component. The first twenty principal components are displayed in the scree plot in Figure 10.

The amount of variation described by the first PC is greatly reduced when compared with the first PC extracted from the variance/covariance matrix. The amount of cumulative variance described by each successive component is also less than that seen in Table 5.8. This change in the description of the variance within the data by the components using the two different matrices can be clearly seen when comparing the two scree plots in Figures 5.3 and 5.10. The absolute values for the eigenvalues are different as the correlation coefficients are re-scaled to remove the effect of peak magnitudes and therefore decreases the absolute value for the eigenvalue. However, a comparison of the slope of the two curves can be made and shows that the slope in the variance/covariance method is somewhat steeper in its descent. This reflects the way in which the absolute magnitudes of the larger peaks dominate the variance measurement within the data.

The correlation coefficient matrix highlights more of the subtle differences between peaks irrespective of their absolute heights. Therefore this method will give a more complete description of the interrelationships between the peaks seen in the chromatographic profile. It does have the drawback, however, that the number of components needed to describe the majority of the variation within the original data set has increased (six PCs to obtain 95% of the variation compared with three PCs with the variance/covariance method).

Table 5.9 *Variance described by first ten components extracted from the matrix of correlation coefficients for FID data for known samples of HFC 134a.*

Principal component	Eigenvalue	Percentage variance	Cumulative variance
1	14.99	40.52	40.52
2	11.03	29.81	70.33
3	4.47	12.09	82.42
4	2.79	7.53	89.95
5	1.18	3.19	93.14
6	0.68	1.83	94.97
7	0.51	1.37	96.34
8	0.43	1.17	97.50
9	0.31	0.84	98.35
10	0.23	0.61	98.96

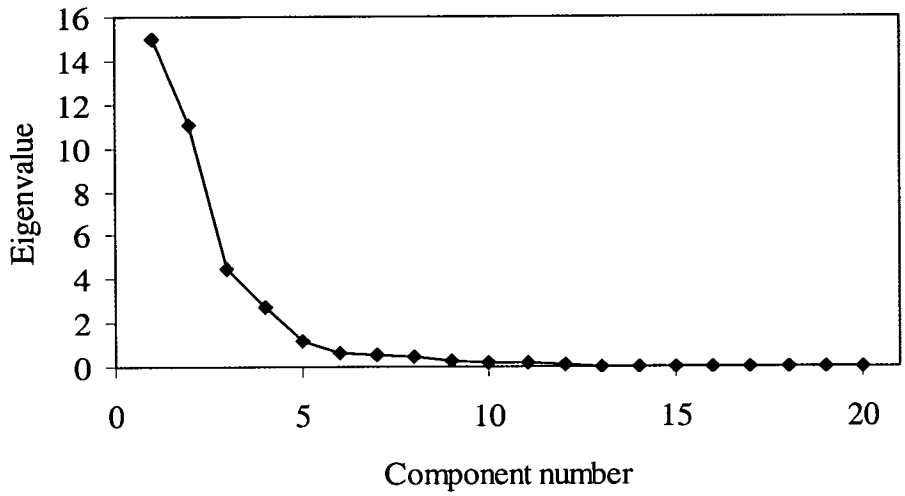
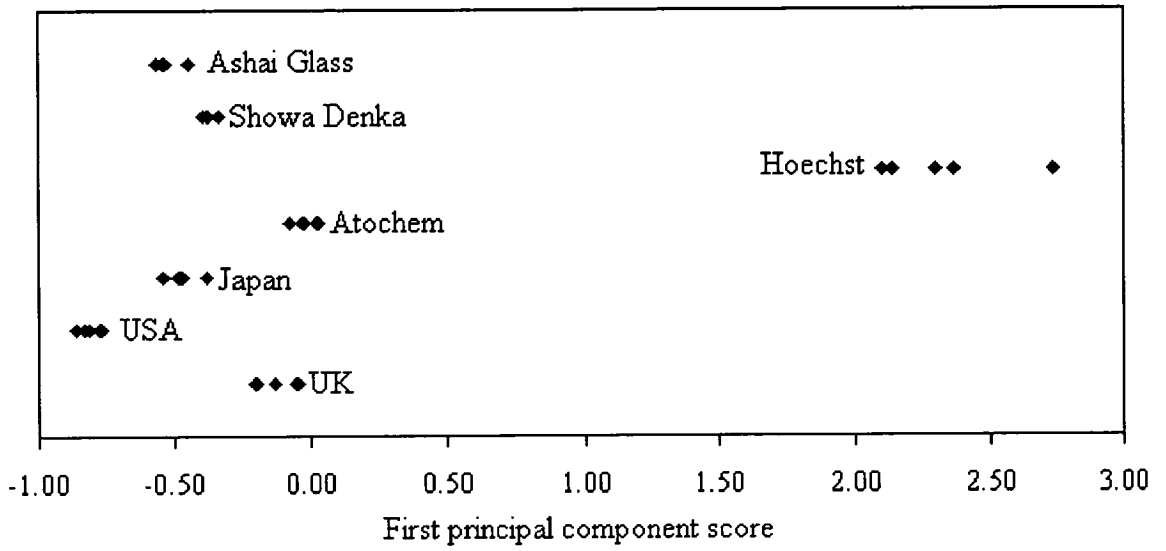


Figure 5.10 *Scree plot for the components extracted from the chromatographic data for known samples of HFC 134a using the correlation coefficient matrix.*

The principal component scores for each sample were calculated and are displayed in Appendix N. The component scores for PC 1 extracted from the correlation matrix are plotted in Figure 5.11 against a standard value. The separation pattern seen in Figure 5.11 is very different from that seen in Figure 5.4 which represents PC 1 extracted from the variance/covariance matrix. PC 1 extracted from the correlation matrix, although it appears to contain less overall variation, allows more of the samples to be separated on the basis of only one variable. The samples representing the Hoechst produced HFC 134a are shown to be distinct from all the other groups. The samples representing Japan, Showa Denka and Ashai Glass have been grouped together on the basis of PC 1 in an overlapping group which is separate from the

other samples. The USA sample has been separated from all the other samples. The UK and Atochem samples are distinct from all the other samples but overlap with each other. Therefore, by looking only at the separation based on PC 1, unknown samples could be placed into one of four groups of which two contain only one



specific sample type.

Figure 5.11 *PC 1 extracted from correlation matrix plotted against an arbitrary value for FID data of known samples of HFC 134a.*

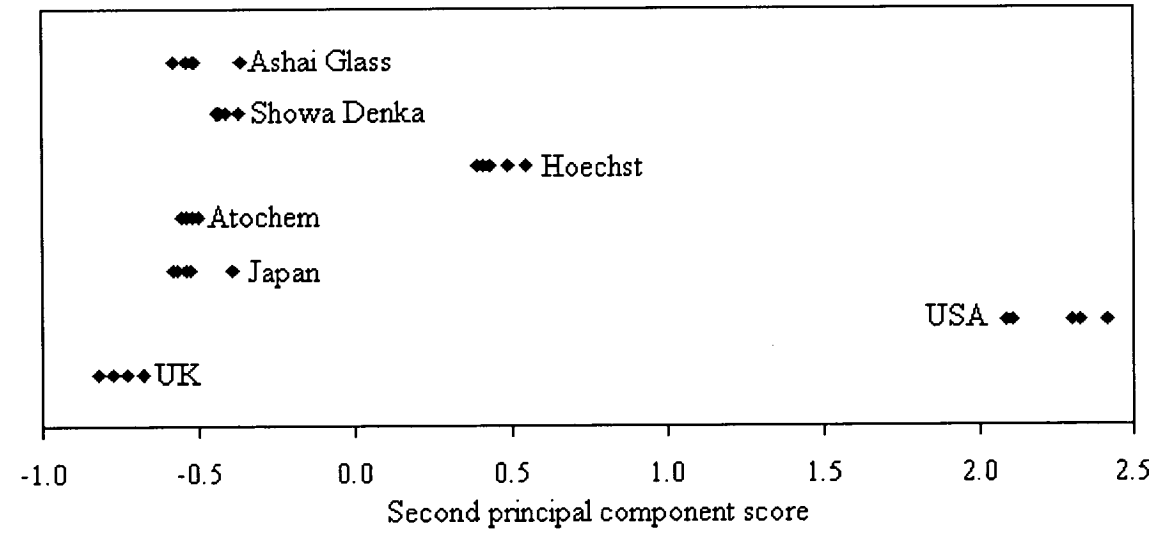


Figure 5.12 *PC 2 extracted from correlation matrix plotted against an arbitrary value for FID data of known samples of HFC 134a.*

PC 2 extracted from the correlation matrix is plotted against an arbitrary value in Figure 5.12. The second principal component, as calculated from the correlation matrix, contains 29.8% of the total variance in the data set. The separation of the

samples on the basis of PC 2 is different from that seen with PC 1. The samples which are now the most different from the other samples are those representing the USA. The samples representing Hoechst are again distinct from all the other samples. The samples which represent the UK and Atochem can now be seen to be distinct from one another, allowing these samples to be distinguished. The samples representing Japan, Atochem, Showa Denka and Ashai Glass cannot be fully distinguished on the basis of PC 2.

The different PCs describe different areas of the variation in the data sets. It does not always follow that the most useful principal components describe the largest amount of variation or the information required is contained within the first few principal components. It is obvious that the information in the first two components does not afford the best separation of the Showa Denka and Ashai Glass samples.

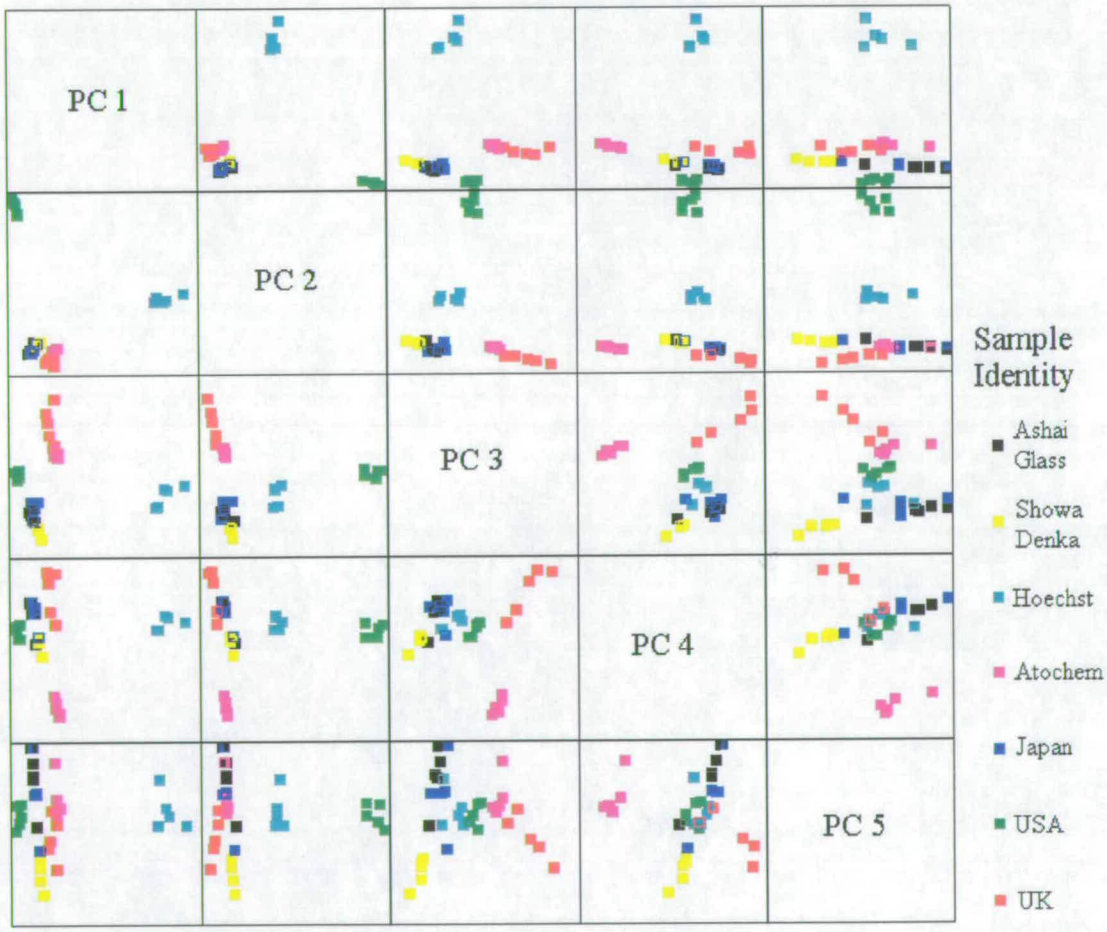


Figure 5.13 *Draftsmans plot of first five principal components extracted from the correlation matrix for the FID data.*

Viewing further principal components allows the identification of which, if any, components are specifically important for separating either a group of samples from the other samples or a specific sub-group. Plotting a Draftsmans plot, see Figure 5.13, allows the comparison of a number of the components (usually less than ten) in an easy to view way which shows which components are important in the separation of the different samples.

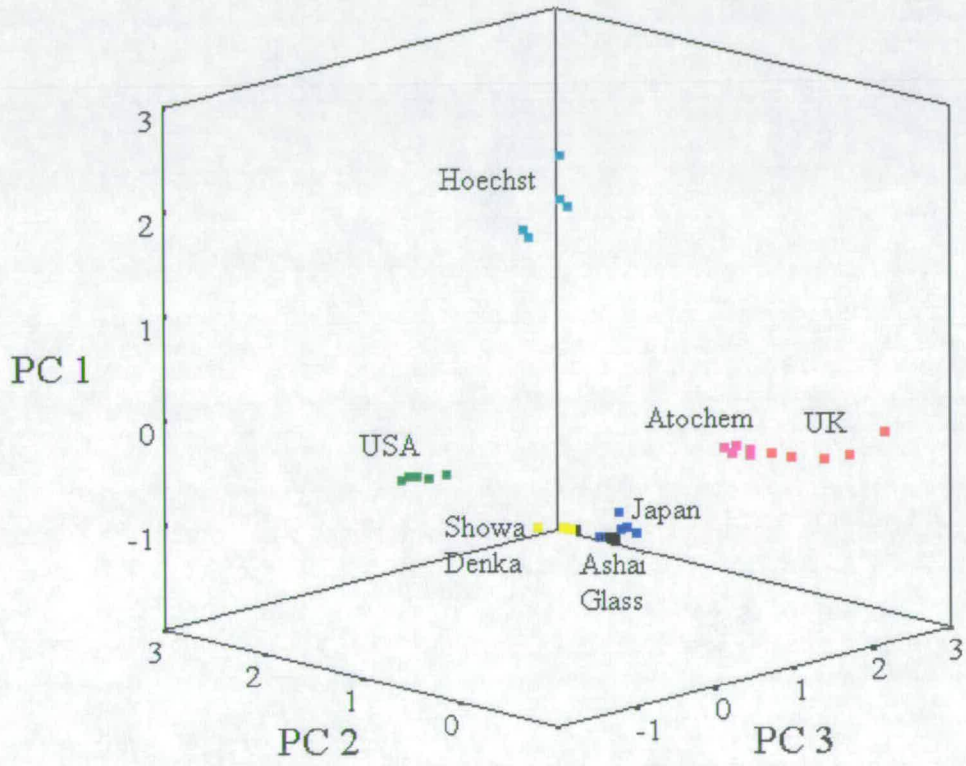


Figure 5.14 *Three dimensional plot for principal component scores 1, 2 and 4 extracted form the correlation matrix plotted against each other for FID data of known samples of HFC 134a.*

Rotation of the axes in three dimensional space (Figure 5.14) can give a better idea of the degree of separation. For example, the samples representing Japan and UK appear close in the 2-D representation, but are in fact well separated in the plane into the page. Reproducing a series of 3-D plots from different angles to show the best view of each separation between the different sample groups can make the representations appear muddled and confusing. Further characterisation of the samples by clustering techniques allows the samples to be distinguished on the basis of formal mathematical rules and will be discussed in Section 6.1.

The PCs extracted from the correlation matrix can be investigated to determine which peaks, hence by-products, have the greatest influence over any of the individual PCs. From this information the relative importance of the different peaks and hence by-products of synthesis in the samples of HFC 134a analysed can be ascertained.

The component matrix contains the values which represent how well each by-product peak is correlated to each PC. That is, if a peak has a value approaching 1 it is highly positively correlated; if the peak has a value approaching –1 it is highly negatively correlated. In both cases, peaks with values near 1 or –1 have major contributions to that specific PC and are therefore important in defining the differences between samples separated by looking at that PC. By-products which have component values approaching 0 are uncorrelated to the PC and therefore have no significant effect on the PC and hence are unconnected to any clustering made on the basis of a PC.

Figure 5.15 shows a bar chart of the correlations between each of the by-product peaks and the first principal component extracted from the correlation coefficients matrix. What is immediately apparent from the figure is that a large number of the by-product peaks are highly correlated to PC 1 and therefore influence the corresponding component scores that each sample has. As the correlation coefficients matrix is not influenced by peak magnitude all the peaks can influence the calculated components to the same extent. In the case of the variance/covariance matrix only the larger peaks had a major effect on the first few components; therefore only gross differences were highlighted.

The correlation coefficients matrix allows the more subtle differences between samples to have an influence on the separation of the samples. This is particularly apparent when trying to separate the Ashai Glass and Showa Denka samples. Both of these samples have characteristically few by-products present and the ones that are present only appear in relatively small amounts. With the variance/covariance matrix

method these peaks effects are 'swamped' by the magnitude of the larger peaks present in other samples.

The object of this project is to use multi-variate information and it is therefore imperative that the method used to investigate the data applies the same weight to the presence of large and small peaks. The final understanding of these differences will take into account the absolute magnitudes of peaks but to perform the initial steps in the separation of samples it is necessary to treat all peaks as of equal importance no matter what their absolute peak height values.

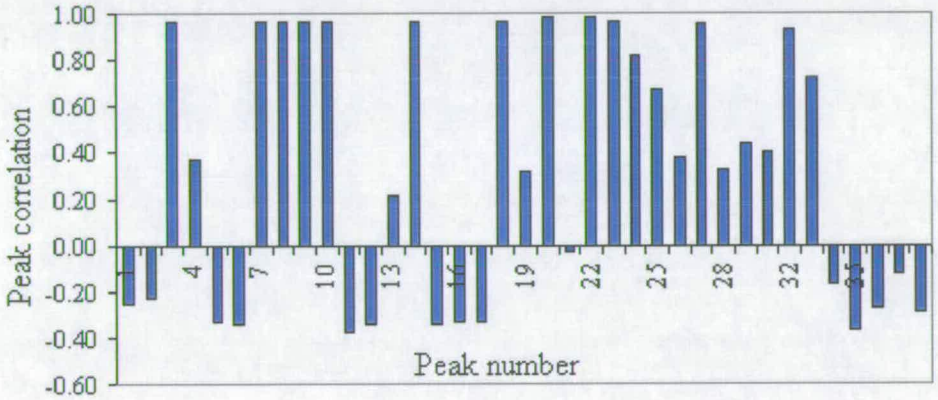


Figure 5.15 *Levels of effect each peak in the FID chromatographic trace has on component 1 extracted from the correlation matrix.*

5.3.2 Extraction of principal components for ECD chromatographic data

The data obtained from the gas chromatography profiling of the known samples of HFC 134a using ECD were entered into an SPSS spreadsheet. The different peaks representing the different by-products detected were entered as the variables and the different samples were entered as the cases. Each cell in the spreadsheet contained the value for peak height seen for any of the by-products in any of the samples analysed. The spreadsheet contained thirty five different cases (samples) which were described by forty six variables (peaks). In addition to the profiles for the thirty five samples of HFC 134a four profiles of air and four profiles of 'pure' helium carrier gas were entered into the spreadsheet to act as controls.

5.3.2.1 PC extraction from the variance/covariance matrix

As discussed in Section 5.3.1.1 extracting PCs from the variance/covariance matrix forms components which are susceptible to variations in the absolute magnitudes of peaks. As it is obvious from the chromatographic profiles formed by ECD analysis of the HFC 134a samples that the peak heights are vastly different there is no need to discuss at length the results obtained from extracting the PCs in this manner.

As an example, PC 1, calculated from the variance/covariance matrix based on the ECD profiling data contains approximately 99% of the variation within the data. The peaks which are most highly correlated with PC1 are in the main those with relatively large peak areas. The component scores for PC1 for each sample show that the Hoechst sample is most different and all the other samples appear clustered together (see Figure 5.16). The effect of the very large peaks seen in the Hoechst samples and how these larger values account for the majority of the variation when the data was unscaled skews the results. The use of the variance/covariance matrix is obviously unsuitable for the analysis of the ECD data due to the widely different peak heights in the ECD chromatographic traces. These vastly different peak heights are an inherent effect of the relative response levels to the different components (*e.g.* those containing chlorine atoms)

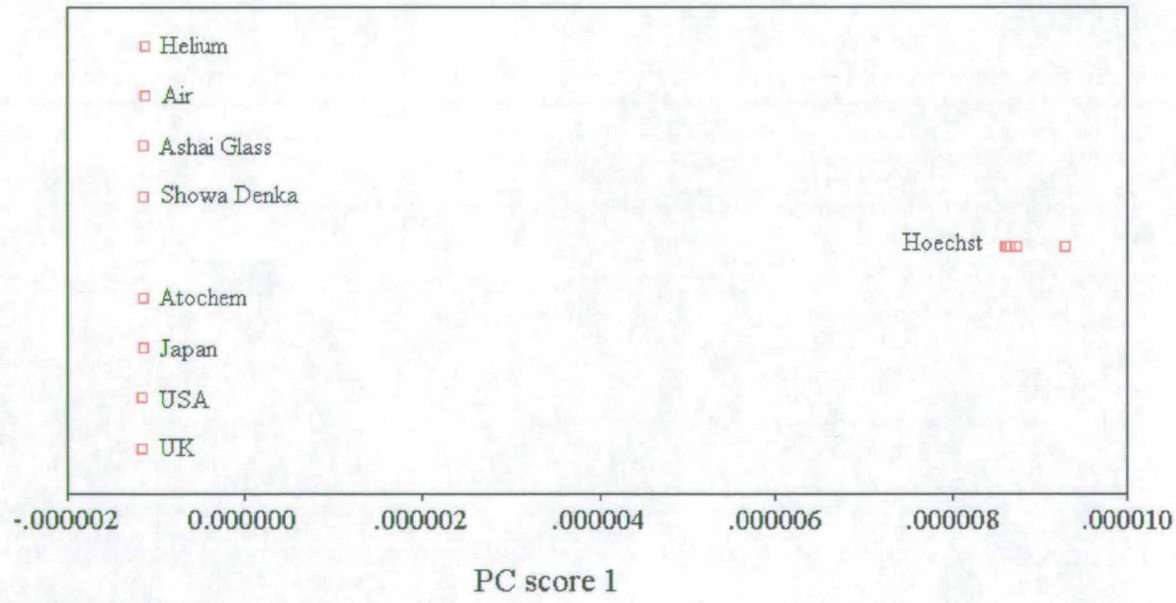


Figure 5.16 *PC 1 scores for data extracted from the variance/covariance matrix for the different samples analysed by ECD.*

5.3.2.2 PC extraction from the correlation matrix

The correlation coefficients matrix was calculated from the data for the ECD profiles of the samples of HFC 134a and blank samples of both air and helium carrier gas. The eigenvalues and eigenvectors were calculated for the correlation coefficients matrix where the eigenvalues were used to calculate the total variance described by each PC. The scree plot representing the amount of variance described by each component is shown in Figure 5.17 with the exact values for the first ten components being displayed in Table 5.10.

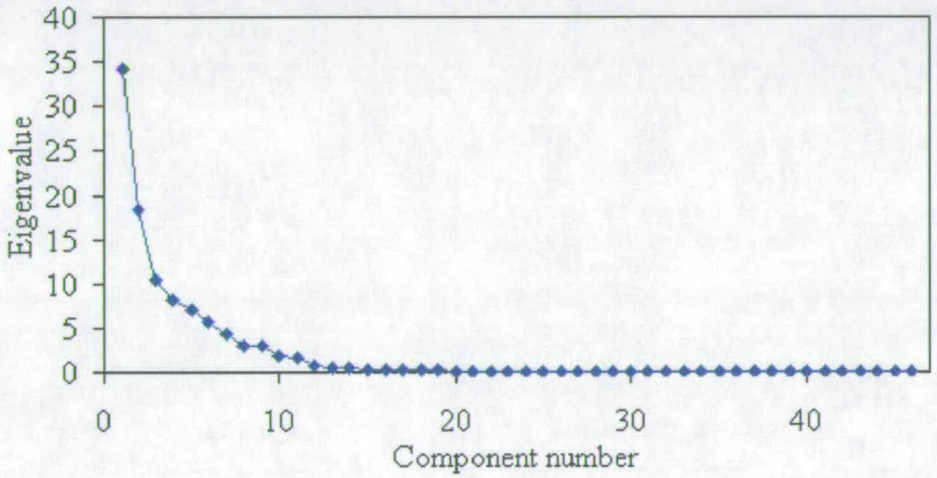


Figure 5.17 *Scree plot for the components extracted from the ECD profiles of HFC 134a samples and blanks using the correlation matrix.*

Table 5.10 *Variance described by first ten components extracted from the matrix of correlation coefficients for ECD data for known samples of HFC 134a and blanks.*

Principal Component	Eigenvalue	Percentage variance	Cumulative variance
1	15.7	34.1	34.1
2	8.4	18.2	52.3
3	4.8	10.4	62.7
4	3.7	8.1	70.8
5	3.2	6.9	77.7
6	2.6	5.6	83.4
7	2.0	4.3	87.7
8	1.4	3.0	90.7
9	1.3	2.9	93.5
10	0.8	1.8	95.4

The variation described by each of the components is more diverse than seen with the FID data based on the correlation coefficients matrix. That is, the first three principal components describe a large degree of the variation within the data (62.7%) but not as much of the variation as seen with the FID data (82.4%).

Considering any one PC in isolation does not allow the full separation of all the sample groups. Figure 5.18 is a Draftmans plot of the first five principal components plotted against each other. This allows the comparison of the samples based on their scores for the first five PCs (see Appendix O for first ten PC sores). However, not all the samples could be clearly separated into distinct clusters on the basis of their origins using this visualisation method.

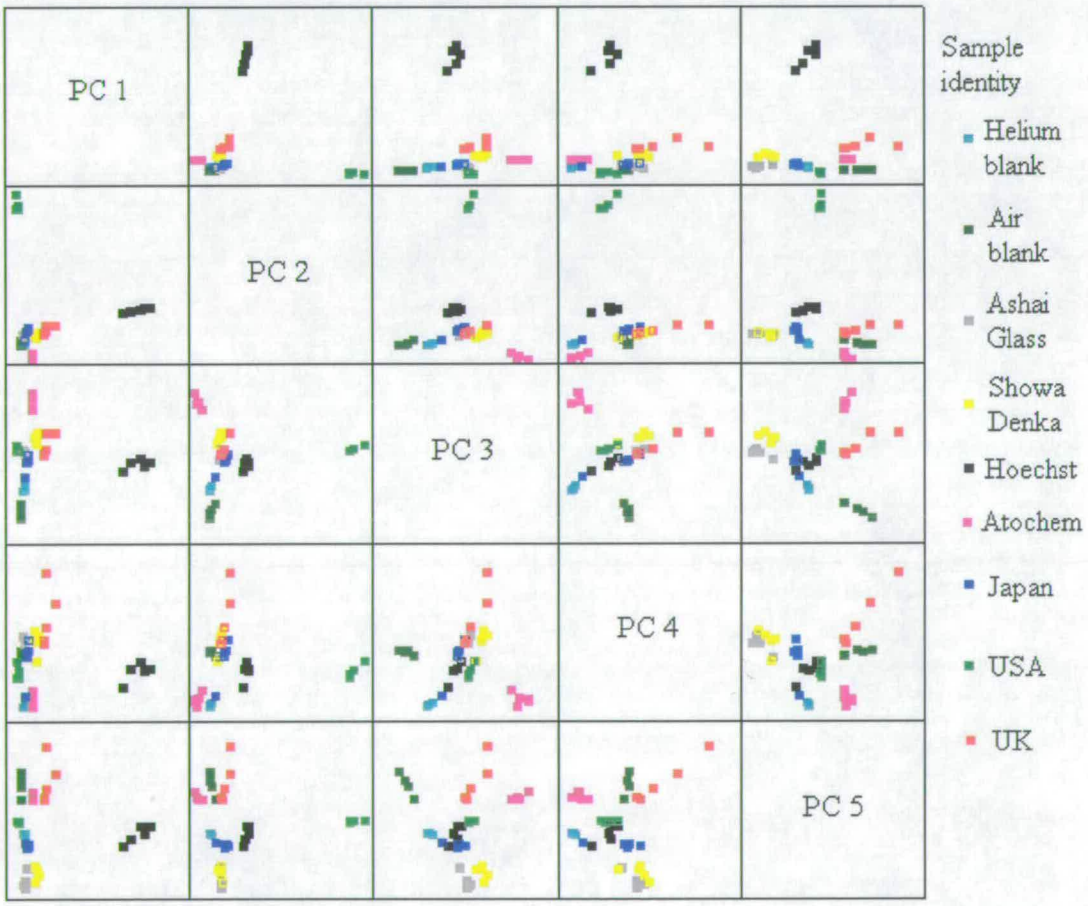


Figure 5.18 *Draftmans plot for the first five components extracted from the correlation matrix for the ECD data.*

Viewing the data points projected onto a 3-D graph does not allow all the samples to be separated into clear groups. A number of samples appear in close proximity and are difficult to distinguish when viewing the data set as a whole.

Considering the data set as a whole it is difficult to distinguish all the samples from one another when using only a limited number of components (<4). However, by considering samples in smaller groups the separations can be seen much more clearly. For example, the Japan and Showa Denka sample groups appear close on the Draftmans plot in Figure 5.18 and appear difficult to distinguish. However, viewing these two groups of samples in isolation by plotting only their component scores in the Draftmans plot produces Figure 5.19. From this figure it can be seen that by viewing the results of the principal component analysis for any two of the first five principal component the ICI Japan samples can be distinguished from the Showa Denka samples.

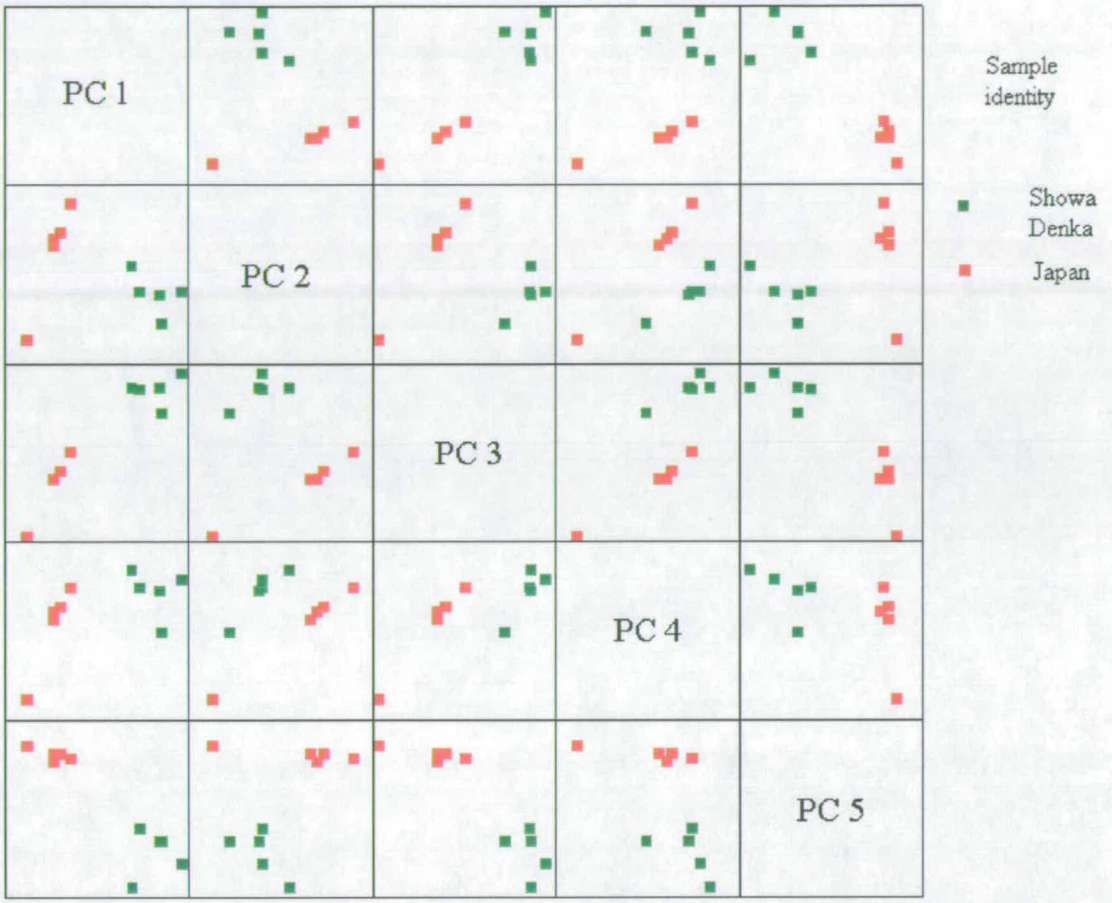


Figure 5.19 *Draftmans plot for ECD results for Japan and Showa Denka samples of HFC 134a in isolation.*

Viewing the data sets in isolated groups, however, reduces the effectiveness of the multi-variate techniques employed as the whole point of the work is to be able to distinguish quickly between samples en masse. The following chapter on cluster analysis (Chapter 6) shows how the data formed in the principal component analysis and the data both raw and transformed can be further investigated without the limitations imposed by trying to visualise the data.

5.3.3 Extraction of PCs from combined chromatographic data

The data obtained from the chromatographic analyses of the HFC 134a samples by FID and ECD have so far been treated separately. The data obtained from the two methods of detection can, however, be treated as one large data set. The order of the peaks/variables is only of interest when identifying the different components present in the samples of HFC 134a analysed. Once the peaks have been identified and

compared between samples they can be treated as distinct variables. That is, by combining the chromatographic profiles of the two detection techniques the amount of information to describe the samples can be increased. This is because different components can be seen using the two different detection methods. Either data set can be added to the other data set to give one large matrix of results. As long as each sample used in the comparison has been analysed by both detection methods all the cases can be compared.

5.3.3.1 Extraction of PCs from correlation coefficient matrix of combined data

The data sets for the FID and ECD analysis of all the known samples of HFC 134a were entered into an SPSS spreadsheet from their corresponding EXCEL spreadsheets. The data for the FID analysis of the samples HFC 134a were entered as the first thirty seven variables. The ECD data were entered for each sample in the corresponding row to the sample analysed by FID and were represented by forty six variables.

The principal components were extracted from the correlation coefficient matrix to remove any weighting effects due to the large differences in relative peak heights. Variable 29 was not used from the FID data set as this peak corresponded to HFC 134a and was off scale in all the samples analysed. Variables 62, 63 and 81 in the mixed matrix of FID and ECD data (peaks 24, 25 and 43 in the ECD data matrix) were not used in the principal component analysis as these peaks were only seen in the air and helium blanks when analysed by ECD. As these two groups of blank samples were not analysed by FID, the ECD data for these samples are not included.

The data matrix for the combined results of the FID and ECD analysis of the thirty five samples of HFC 134a described by a total of eighty peaks was subjected to a principal component analysis of the correlation coefficient matrix. Table 5.11 shows the eigenvalues and the percentage variation of the first ten principal components. Figure 5.20 is the scree plot of the first twenty principal components. The total

variance described by the first ten principal components is 95.7 % of the original variance within the two data sets.

Table 5.11 *Variance described by the first ten principal components extracted from the matrix of correlation coefficients for combined set of FID and ECD data for known samples of HFC 134a.*

Principal Component	Eigenvalue	Percentage variance	Cumulative variance
1	30.6	38.3	38.3
2	19.2	24.0	62.2
3	7.9	9.9	72.1
4	6.8	8.5	80.7
5	3.5	4.3	85.0
6	2.5	3.2	88.2
7	1.7	2.2	90.4
8	1.6	2.0	92.4
9	1.5	1.8	94.2
10	1.2	1.5	95.7

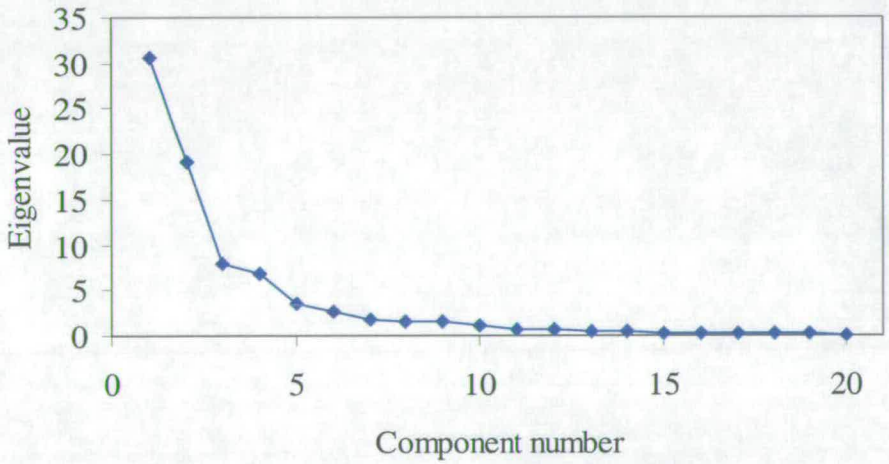


Figure 5.20 *Scree plot for the first 20 components extracted from the combined chromatographic data for the FID and ECD analysis of HFC 134a samples from known origins.*

The component scores for each of the samples were calculated for the first 10 PC and are displayed in Appendix P. The PC scores for PC 1 are plotted against a set value in Figures 5.21a and b, where 5.21b is an enlarged scale of 5.21a on which the Hoechst samples do not appear. Figure 5.21a clearly shows that the Hoechst samples and USA samples can both be distinguished from all other samples analysed. The

enlarged scale of Figure 5.21b shows that the samples representing Japan and Ashai Glass can be grouped separately from the other samples but not distinguished from each other.

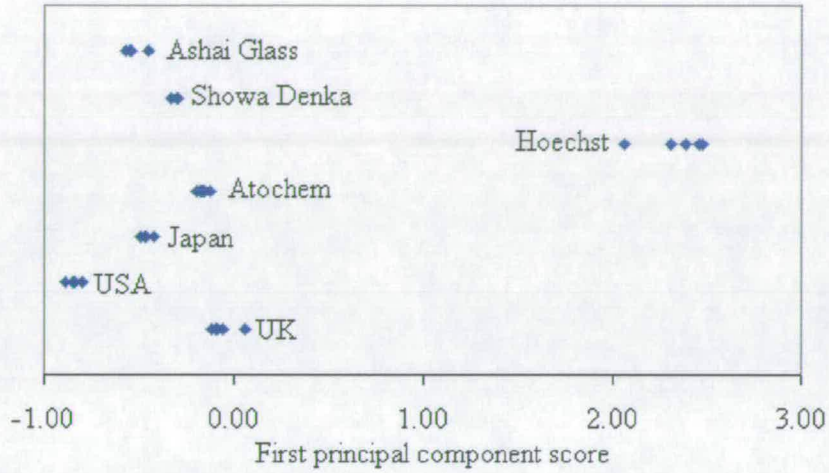


Figure 5.21a *PC 1 extracted from the correlation coefficients matrix plotted against an arbitrary value for mixed data matrix of FID and ECD results for known samples of HFC 134a.*

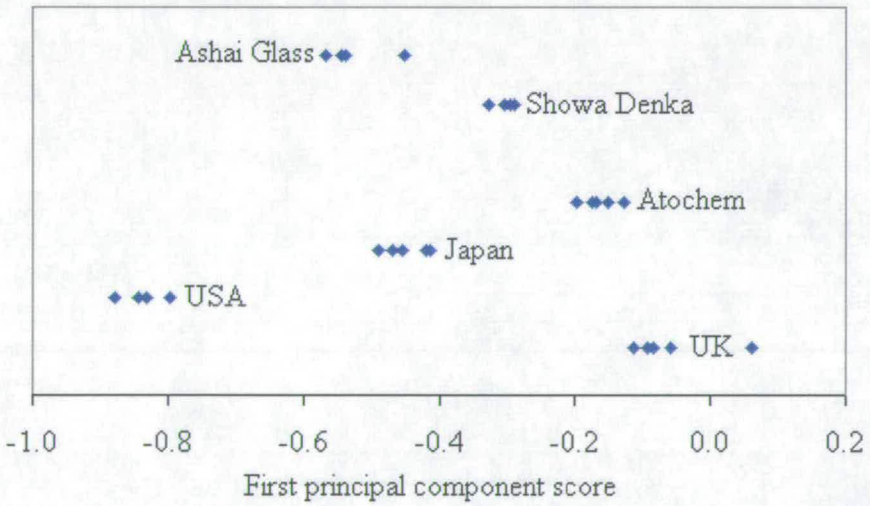


Figure 5.21b *PC 1 extracted from the correlation coefficients matrix plotted on an enlarged scale against an arbitrary value for mixed data matrix of FID and ECD results for known samples of HFC 134a.*

The samples representing Showa Denka are separated from all the other samples into a distinct group. The samples representing UK and Atochem do not overlap with each other and can therefore be described as being separate groups although the two closer values in each group (-0.11 for UK and –0.13 for Atochem) do not show a high degree of separation.

The separation of the samples of HFC 134a from known origins is not possible by looking at PC 1 in isolation as the sample groups for Japan and Ashai Glass overlap. However, the samples only overlap because of one sample from the Ashai Glass group (the first sample in the group). By looking at the correlation between the original peaks and PC1 it is possible to see which peaks are having the majority of effect on the formation of PC 1. By ascertaining which peaks are important in the separation of the samples it is possible to return to the original data and see specifically which peak values are causing the overlap in the two sample groups.

Looking at Figure 5.22 it is obvious that there are a large number of peaks which are highly positively correlated to PC 1 (values approaching 1). However, the original data set of peaks can be reduced to those peaks which are contained in either or both Japan and Ashai Glass *i.e.* if both the Japan and Ashai Glass sample do not contain a specific peak then this peak is irrelevant when comparing these two samples in isolation. The removal of missing peaks in both samples reduces the number of variables to forty one. The second step is to see which of these remaining peaks are highly correlated with PC 1. Of the forty one peaks sixteen of these show high correlation with PC1 (values >[0.7]).

Looking at the peak height values in the raw data for the samples from Japan and Ashai Glass (Appendix G and H) it is possible to see the differences the PC analysis is highlighting. The outlying Ashai Glass sample has an overall trend of lower peak heights than the other samples. When looking at peaks 53, 64, 71 and 82 this brings the value of the Ashai sample closer to the peak heights for the Japan samples. It is this general trend which leads to the grouping of the outlying sample of Ashai Glass with the Japan samples.

To distinguish between the samples of Japan and Ashai Glass more components need to be employed. In particular, those components which show high correlations to the peaks in the original data that highlight the difference between the samples *i.e.* peaks 30 and 31 where the Ashai Glass samples have higher values than the Japan samples.

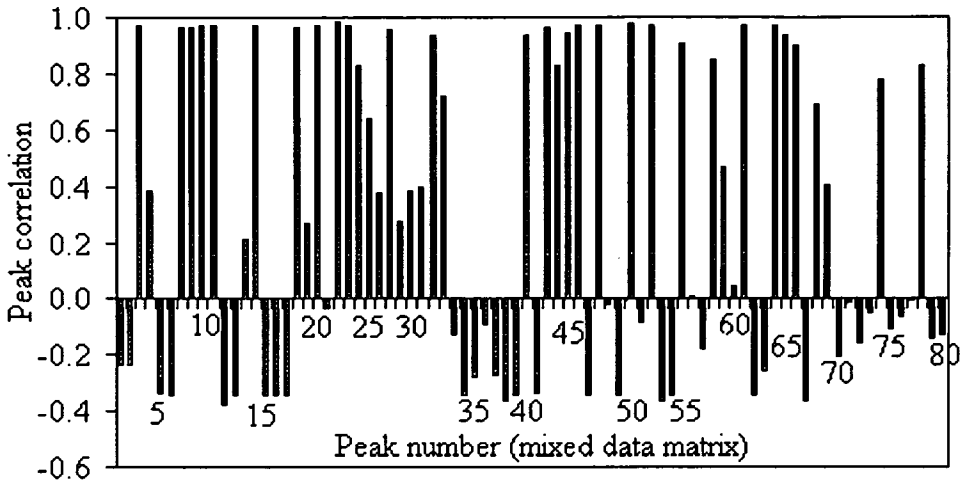


Figure 5.22 *Degree of correlation between each peak in mixed matrix of data for FID and ECD of known samples of HFC 134a and PC 1 extracted for correlation coefficients matrix.*

Also, as the samples were distinguished completely, apart from one outlying Ashai Glass sample, any component which separates this sample from the Japan group would allow the differentiation of the two groups. The second component does not necessarily need to group the outlying Ashai Glass sample with the other Ashai Glass samples as long as it separates it from the Japan group. Figure 5.23 shows the separation of the two sets of samples when looking at PC 5 and PC 7. In both sets of samples there appears to be an outlier as two samples are not clustered with the other values. However, as the origins of the samples are known it is possible to draw a discriminant line between the two sets of samples.

The separation of the two sets of samples from each other was best achieved when looking at PC 5 and PC 7. The problem with using all the data from both detection techniques and all the samples analysed is that some of the subtle differences between the samples may be described by a number of components which when viewed individually or as pairs are lost. That is, variation which would allow the separation of the two sets of samples may be distributed amongst a number of different components which are dominated by the description of other peak variations. Therefore subtle differences between samples can be swamped by the gross effect of differences within other samples.

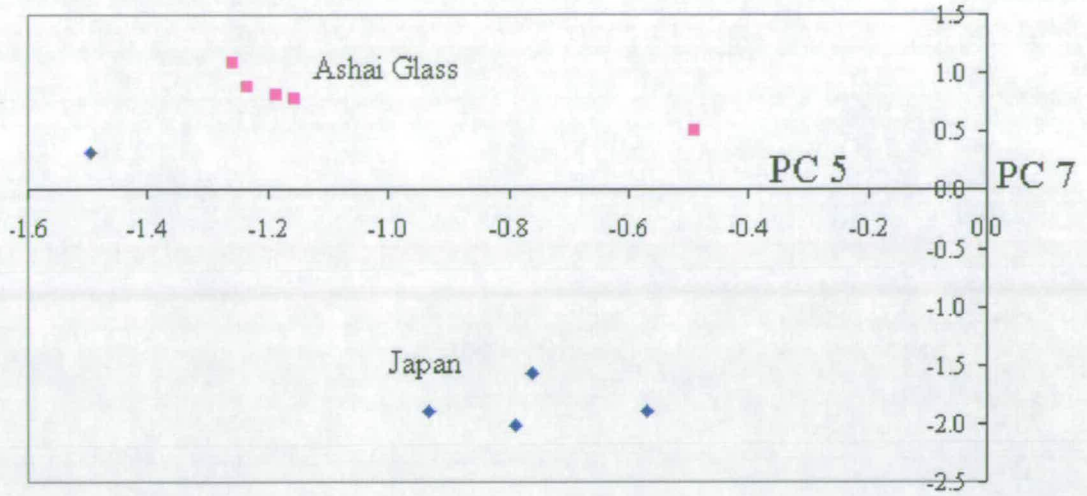


Figure 5.23 *Separation of Ashai Glass and Japan samples on the basis of PC5 and PC7 extracted from the combined correlation coefficient data matrix of FID and ECD results.*

This effect can be illustrated by extracting the principal components from the correlation coefficients data matrix corresponding to just the Japan and Ashai Glass samples. Plotting PC 1 against a set value gives the results in Figure 5.24. From this figure it is plain that the two groups of samples are clearly distinguishable from one another. As the results from the PC analysis of all the data sets just represent the variance in a different way it is only an interpretation problem that prevents the grouping of the different samples.

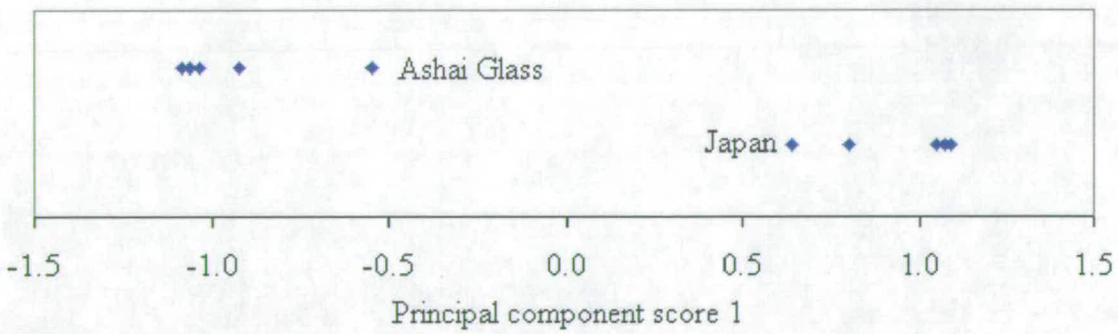


Figure 5.24 *Separation of Ashai Glass and Japan samples on the basis of PC1 extracted from the combined correlation coefficient data matrix of FID and ECD result for only the Japan and Ashai Glass samples.*

The following chapter details the mathematical classification methods employed to overcome the limitations encountered when visually inspecting larger data sets containing multiple groups[13].

References

[1] Mead, R., Curnow, R.N. and Hasted, A.M., *Statistical Methods in Agriculture and Experimental Biology*, 2nd ed., Chapman and Hall, London, p317 (1993)

[2] Kendall, M.G and Buckland, W., *A Dictionary of statistical Terms*, 4th ed., Longman Group, London, (1982)

[3] Motulsky, H.J., Stannard, P. and Neubig, R., *GraphPad Prism™*, Version 2, San Diego, p169-171, (1995)

[4] Adams, M.J., *Chemometrics in Analytical Spectroscopy*, Royal Society of Chemistry, Cambridge, (1995)

[5] Fales, H.M. and Jones, T.H., *Anal. Chem.*, **52**, p982-983, (1980)

[6] Box, G.E.P., and Cox, D.R., *J. R. Statist. Soc.*, **B36**, p211, (1964)

[7] Lamparczyk, H., Ochocka, R.J., Grzybowski, J., Halkiewicz, J. and Radecki, A., *Oil and Chemical pollution*, **6**, p177-193, (1990)

[8] Bicchi, C.P., Panero, O.M., Pellegrino, G.M. and Vanni, A.C., *J. Agric. Food chem.*, **45**, p4680-4686, (1997)

[9] Manly, B.F.J., Manly, *Multivariate Statistical Analysis: A Primer*, Chapman and Hall, London, UK, (1991)

[10] Wold, S., *Pattern Recognition*, **8**, p127-139, (1976)

[11] Wold, S., *Chemom. Intell. Lab. Syst.*, **2**, p37-52, (1987)

[12] Hopke, P.J., *Chemomet. Intell. Lab. Systems*, **6**, p7, (1989)

[13] Stout, S.A., *J. of Anal. Appl. Pyrolysis*, **18**, p277-292, (1991)

Chapter 6

Chemometric Data Analysis

6.0 Introduction to chapter

The following chapter details the work performed to classify the different samples of HFC 134a based on chemometric techniques. The chapter details how the two unsupervised classification techniques of K-means clustering and hierarchical clustering were used to attempt to separate the samples into distinct groups, where each group represents a set of samples from a specific production origin. These techniques were applied to the raw data, the normalised data and the PC scores for both the FID and ECD data. The clustering techniques and data formats which provided the best separation, where the best separation is defined as placing each sample into a cluster/group populated by only samples from the same origin of production, were taken and used to classify unknown samples in Chapter 7.

6.1 Unsupervised cluster analysis of known samples

The classification of the different samples of HFC 134a was performed using a number of cluster analysis techniques applied to each of the data sets described in Chapter 5. The two techniques of K-means clustering (see Section 2.2.4) and hierarchical clustering (see Section 2.2.3) have both been widely used in multi-variate classifications[1,2]. These were applied to the following data sets extracted from the FID chromatograms of the known samples: the raw data, the scaled data and the principal component scores calculated from the correlation coefficient matrix. The two clustering techniques were also applied to the data sets extracted from the ECD chromatograms of the known samples for the raw data, the scaled data and the principal component scores from the correlation coefficient matrix. Finally, the clustering techniques were applied to the FID and ECD combined data sets of raw, scaled and principal component scores for the known samples of HFC 134a. The success of the clustering techniques was measured by the number of misclassifications of the known samples, with the ultimate goal of attaining no misclassifications *i.e.* all samples placed in clusters populated only by samples from the same production origin.

6.1.1 Classification of known samples using K-means clustering

The three data sets of raw, normalised and principal component scores were each analysed by K-means clustering. K-means clustering can be applied using a number of different cluster centre calculations in order to tailor the results to produce the best separation for the data set analysed and to avoid local minima in the iterative process[3]. For each analysis of the different data sets the number of cluster centres which provided the best separation of the data set are presented and discussed.

6.1.1.1 Classification of raw FID data using K-means clustering

The FID raw data set consisted of thirty five samples of HFC 134a from seven different known sources. These samples were described by thirty seven variables representing the thirty seven different peaks seen in the FID chromatographic traces. The values for these peaks ranged from not seen (0.015mV) to the largest peak at 43.854mV (USA sample 3, peak 13). The initial cluster analysis was performed using seven cluster centres to define the samples as this was the expected number of groups to be seen.

The results of the K-means clustering of the known samples of HFC 134a analysed by FID are displayed in Table 6.1. The clustering technique has correctly separated the Atochem and Hoechst samples into two clusters distinct from all other samples of HFC 134a. However, the technique has not distinguished between the Japan, Showa Denka and Ashai Glass samples and has placed them into one large group. The two sets of samples from the UK and USA have been separated into four groups. From these results it can be concluded that the differences within some of the groups of samples (*i.e.* UK or USA) as calculated using this technique are greater than the differences between some of the samples (*i.e.* Japan and Showa Denka).

Table 6.1 *Results of K-means cluster analysis of FID raw data.*

Sample identity	Cluster	Distance from cluster centre	Sample identity	Cluster	Distance from cluster centre
UK 1	3	1.29	Japan 1	5	3.37
UK 3	3	1.65	Japan 2	5	2.35
UK 5	3	1.54	Japan 3	5	2.86
UK 2	4	1.82	Japan 4	5	1.61
UK 4	4	1.82	Japan 5	5	1.63
USA 1	6	2.01	Showa Denka 1	5	2.81
USA 2	6	3.16	Showa Denka 2	5	1.80
USA 3	6	2.08	Showa Denka 3	5	1.72
USA 4	1	1.06	Showa Denka 4	5	1.62
USA 5	1	1.06	Showa Denka 5	5	1.59
Atochem 1	2	1.41	Ashai Glass 1	5	2.28
Atochem 2	2	0.58	Ashai Glass 2	5	1.59
Atochem 3	2	1.61	Ashai Glass 3	5	2.33
Atochem 4	2	1.50	Ashai Glass 4	5	1.74
Atochem 5	2	0.74	Ashai Glass 5	5	1.54
Hoechst 1	7	4.89			
Hoechst 2	7	3.83			
Hoechst 3	7	3.85			
Hoechst 4	7	3.45			
Hoechst 5	7	3.08			

The classification of the samples on the basis of the raw data is prone to effects of magnitude, therefore careful consideration should be given to pre-treatment of the data before classification analysis[4]. The percentage variation within a group of samples for any given peak has been shown to be relatively consistent (between 0-20%, see Section 5.1.2); however, the absolute variations can be relatively large. For example, Table 6.2 shows the peak heights for peak 13 for the two sets of samples, USA and Showa Denka. Figure 6.1 is a graphical representation on a linear scale of the peak height for peak 13 for the samples of USA and Showa Denka. From the table, the values for the USA samples can be seen to be approximately two orders of magnitude greater than the Showa Denka samples. The absolute difference between the largest and smallest values in the USA samples is approximately fifty times greater than with the Showa Denka samples. Viewing these results on the graph (Figure 6.1) it appears the Showa Denka samples are more precise and if a K-means

cluster analysis was performed on this single variable all the USA samples would be placed into different groups before the Showa Denka samples are separated. However, the variation within the sets of samples is larger within the Showa Denka samples than within the USA samples as shown by the percentage difference (18% and 14.5%). Therefore this method of classification is of limited use when using the raw data which have large variations between the peak heights of the samples studied.

Table 6.2 *Data for peak height of Peak 13 for both USA and Showa Denka samples.*

Sample identity	Peak 13 (mV)	Absolute difference (mV)	Average (mV)	Percentage difference
USA 1	43.67	5.95	41.09	14.5
USA 2	41.65			
USA 3	43.85			
USA 4	37.91			
USA 5	38.35			
Showa Denka 1	0.654	0.11	0.587	18.0
Showa Denka 2	0.593			
Showa Denka 3	0.580			
Showa Denka 4	0.562			
Showa Denka 5	0.548			

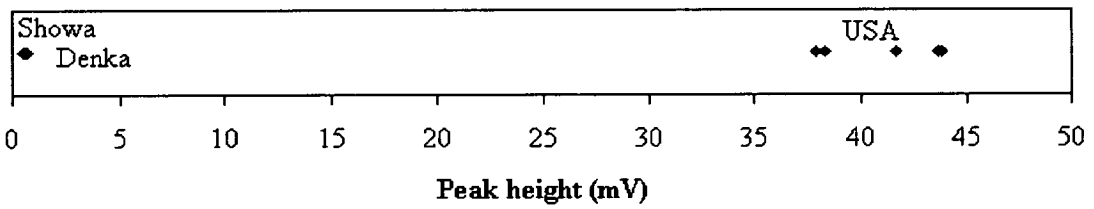


Figure 6.1 *Linear plot of peak 13 height for USA and Showa Denka samples*

6.1.1.2 Classification of normalised FID data using K-means clustering

The FID normalised data set consisted of thirty five samples of HFC 134a from seven different known sources. These samples were described by thirty seven variables representing the thirty seven different peaks seen in the FID chromatographic traces (minus the peak for HFC 134a). The values for these peaks ranged from below LOD (0.015mV/largest peak height) to the largest value of any peak, 1.0 (because of the normalisation). The table of transformed data has a very different variation pattern.

The values for each peak, although scaled to a maximum of 1 still contain the same amount of variation. This can be shown by calculating the precision of the values for any peak (see Equation 6.1). These precision values are the same for the raw data as for the normalised data *i.e.* normalising the data does not affect the results within a peak. However, the variation across the peaks representing any one sample is vastly reduced.

$$precision = \left(\frac{std.dev.}{average} \right) * 100 \qquad \textbf{Equation 6.1}$$

Figure 6.2 shows the results of calculating the precision values for Peaks 1, 2, 37 and 38 for the raw and normalised data. The value for the precision of each peak (reading down the table) is the same in both the raw and normalised data (18.5, 202, 39.8 and 136%, respectively). The figure also shows that the calculation of the precision across the peaks (*i.e.* across a sample) is reduced when using the normalised data. From this it can be concluded that normalising the data has no effect on the intra peak variation between samples but reduces the inter peak variation between samples. This removal of the inter peak variation is because of a reduction in the effect of the absolute values of the peaks.

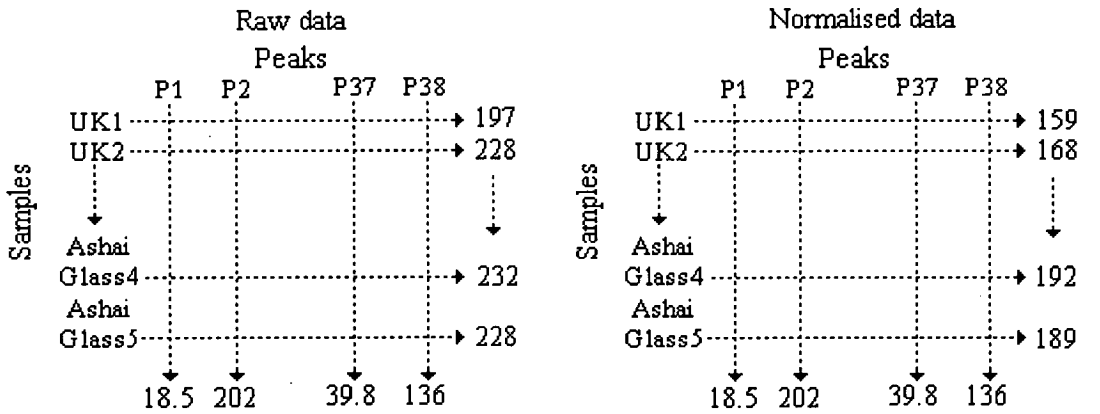


Figure 6.2 *Difference in precision calculation results for raw and normalised FID chromatographic data(expressed as %RSD).*

The K-means clustering was performed using seven cluster centres to represent the expected seven groups of known samples. The initial cluster membership obtained with this number of cluster centres led to the misclassification of a number of the

samples. However, two of the misclassified samples were the furthest distance from the cluster centre to which they were assigned. Therefore, a second cluster analysis was performed using eight cluster centres to describe the data. This data pattern placed the two misclassified samples into a distinct group but did not improve the level of misclassification of samples. Further increases in the number of cluster centres did not improve the cluster assignment of samples.

The cluster membership for eight cluster centres is shown in Table 6.3. The samples representing the UK, USA, Atochem and Showa Denka have all been split into distinct groups representing their origin of production. The Hoechst samples have been split into two separate groups. The majority of the Japan and Ashai Glass samples have not been distinguished and have been placed into one group. Japan 2 and Ashai Glass 1 samples have been placed into a separate cluster. This cluster represents the group formed by increasing the number of clusters from seven to eight.

Table 6.3 *Cluster membership based on K-means clustering of normalised FID chromatographic data with eight cluster centres.*

Sample identity	Cluster	Distance from cluster centre	Sample identity	Cluster	Distance from cluster centre
UK 1	1	0.49	Japan 1	7	0.29
UK 2	1	0.63	Japan 3	7	0.46
UK 3	1	0.41	Japan 4	7	0.28
UK 4	1	0.41	Japan 5	7	0.33
UK 5	1	0.29	Ashai Glass 2	7	0.20
USA 1	5	0.63	Ashai Glass 3	7	0.37
USA 2	5	0.59	Ashai Glass 4	7	0.23
USA 3	5	0.49	Ashai Glass 5	7	0.17
USA 4	5	0.44	Showa Denka 1	3	0.23
USA 5	5	0.46	Showa Denka 2	3	0.07
Atochem 1	4	0.27	Showa Denka 3	3	0.09
Atochem 2	4	0.15	Showa Denka 4	3	0.12
Atochem 3	4	0.18	Showa Denka 5	3	0.13
Atochem 4	4	0.13	Japan 2	8	0.27
Atochem 5	4	0.30	Ashai Glass 1	8	0.27
Hoechst 1	2	0.57	Hoechst 2	6	0.30
Hoechst 3	2	0.32	Hoechst 4	6	0.30
Hoechst 5	2	0.35			

The cluster membership obtained from performing the K-means clustering on the normalised data produces a better separation of the samples than clustering using the raw data set, where the best separation is defined as the number of samples placed into their correct distinct clusters according to their origin of production. The results of the clustering can be displayed as two Venn diagrams[5] to allow easy comparison of the two methods (see Figure 6.3). Note that increasing the number of cluster centres to eight in the raw data analysis led to the splitting of the Hoechst sample into two groups.

The Venn diagrams show how each method clusters the different samples of HFC 134a. Those clusters which contain samples from more than one different origin are highlighted in grey. Samples from the same production origin that have been placed in separate clusters are not considered as misclassified as these clusters are distinct from the other samples. However, it is obviously more desirable to be able to define each cluster to represent a single production origin. The use of the raw data has misclassified/been unable to separate fifteen samples which correspond to all the Japan, Showa Denka and Ashai Glass samples. Further, the UK, USA and Hoechst samples have been split into two clusters. The normalised data have only misclassified/been unable to separate ten samples, which represent the Japan and Ashai Glass groups. This use of the normalised data has also only split one of the sample groups into two distinct clusters.

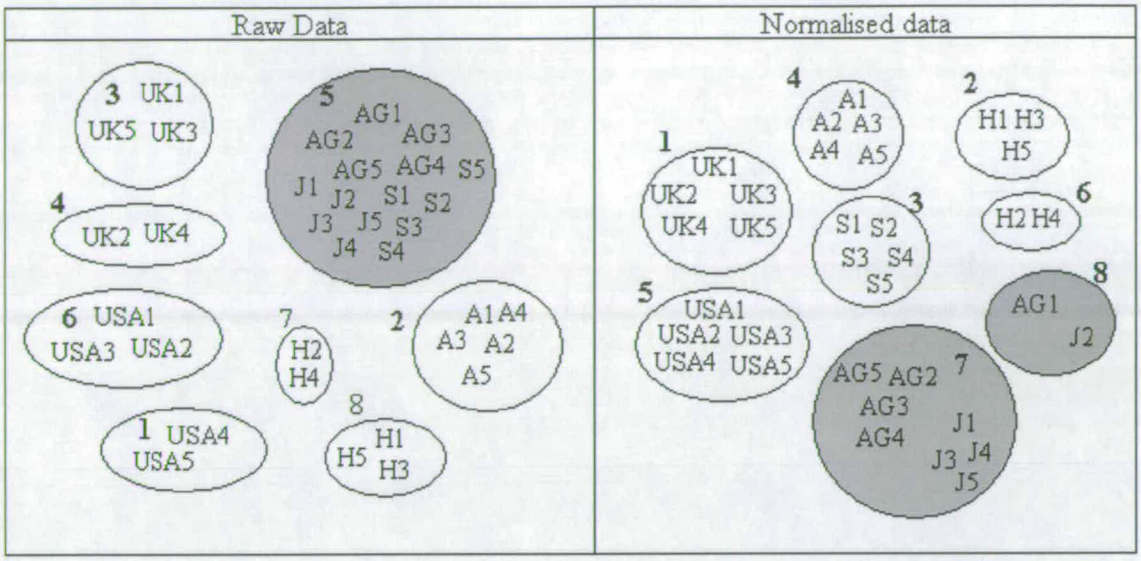


Figure 6.3 *Venn diagram of cluster patterns formed from K-means clustering performed on raw and normalised FID data using eight cluster centres where samples are labelled UK=UK, USA=USA, J=Japan, A=Atochem, H=Hoechst, S=Showa Denka and AG=Ashai Glass*

6.1.1.3 Classification of PC scores for FID data using K-means clustering

The data set representing the PC scores for the FID data was considered in a number of different ways as the number of variables which were used to describe the data is not a set value. The PC1 scores for the FID data may provide enough information for the separation of the data into distinct groups on the basis of their origins. The technique of K-means clustering can therefore be applied to varying numbers of PCs in order to achieve the best separation of groups using the fewest variables[6]. The cluster analysis results for the FID data using various numbers of PCs to describe the data are shown in Table 6.4

Table 6.4 *K-means cluster analysis of FID data using seven cluster centres with a range of PC scores calculated from the correlation coefficient matrix.*

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Correct cluster	20	10	10	20	20	20	14	10	10	0
Distinct cluster	5	10	10	5	5	5	7	10	10	8
Misclassified	10	15	15	10	10	10	14	15	15	27

A sample was classed as correctly clustered if it met the following criteria:

- it was in a group which contained a minimum of four samples from one origin and no more than one sample from any other origin. That is, a cluster which contained four samples from the UK, one sample from the USA and one sample from Atochem contains four samples correctly clustered (UK) and two samples misclassified (USA and Atochem). However, if the cluster contained four samples from the UK and two samples from the USA, the samples were considered not to have been separated and therefore all the samples were considered as misclassified.

A sample was classified as distinctly clustered if it met the following criteria:

- a group of samples (less than four, including single samples) from the same origin were placed in a distinct cluster which contained no other samples. For example, a cluster containing only two samples from one origin and no other samples from any other origin and still be considered as a distinct cluster.

A sample was classified as misclassified it did not meet either of the above two criteria.

Table 6.4 shows that the best classification of the samples was achieved using four, five or six PCs. Using any of these optimum numbers of PC scores to describe the data allowed twenty samples to be correctly classified, ten samples to be distinctly classified and ten samples to be misclassified/not separated (Japan and Ashai Glass samples in all three instances). Using fewer than four PCs led to a decrease in the number of samples correctly classified and an increase in both the numbers distinctly classified and misclassified/not separated. Increases to over six in the number of PCs used led to a deterioration in the classification of the samples.

It is apparent, therefore, that there is an optimum number of PCs required to produce the best classification when using K-means clustering. Classification using fewer than the optimum number of PCs means that not enough of the original variation is considered to allow clear separation of the samples[7] (<4 PCs describes only 82.4%).

Classification of the samples using more than the optimum number begins to use PCs describing random variation between the samples and therefore supplies no useful information for classification purposes.

6.1.1.4 Classification of FID data using a two step K-means clustering

The previous three sections have shown that by using K-means clustering the known samples of HFC 134a cannot be clearly separated into their seven respective origins of production. However, as these techniques rely on measuring variation they are subject to a number of constraints. Consider Figure 6.3 (normalised results) and the results for the K-means cluster analysis based on PC 1 alone (Appendix N, column PC 1). From the figure it could be concluded that the Hoechst samples form a distinct group which is highly different from all the other samples. However, the K-means cluster analysis results using PC 1 splits the Hoechst samples into three separate groups. The results are mathematically correct as in doing this the total distances between the samples and the cluster centres are minimised, but obviously the clustering is not applicable in the context of this work. The effect of the intra-sample variation is to mask the effect of the inter-sample variation in some cases. This is particularly true in the samples which have a limited number of peaks which are describing them. It is therefore possible to distinguish first on the basis of the gross differences between the samples before repeating the analysis using the samples which have been misclassified, in order to separate the samples on the basis of the more subtle differences.

The re-classification of the Japan, Showa Denka and Ashai Glass samples on the basis of their raw data results was performed using three-six cluster centres for the K-means analysis. No re-clustering model separated the samples into their correct origins of synthesis; however, the four cluster model distinctly classified the Japan samples and correctly clustered all but one sample from Ashai Glass. The re-classification of the normalised data using three-six cluster centres for the K-means cluster analysis also failed to correctly cluster all the samples. The best classification was achieved with both five and six cluster centres which correctly

classified nine samples, distinctly classified four samples and misclassified two samples.

The re-classification of the Japan, Showa Denka and Ashai Glass samples on the basis of their PC scores calculated from the correlation coefficients matrix of all the FID data set was performed using the first five PCs. Varying the number of cluster centres from three to six did not produce a clear classification of the samples. However, by taking the raw data representing the samples from Japan, Showa Denka and Ashai Glass and subjecting this smaller data set to a PCA a new set of PC scores can be calculated from the correlation coefficients. This second set of PC scores was then subjected to K-means cluster analysis using only the first two PCs and three cluster centres. The results for this analysis correctly separated the three sets of samples into three distinct groups on the basis of their origins of production.

Using a two step process based on the PC scores of all the data set (data for all samples analysed by FID) followed by a second PCA of the sub-set of the data containing only those samples misclassified, it was possible to separate the samples on the basis of their FID chromatographic profiles. A second separate PCA of the sub-set of samples is required because of the quality of the data obtained from the FID analysis of these samples. The number of peaks which were present in these samples was relatively small (eighteen peaks compared to thirty seven peaks). This means that these samples were classified more on the basis of the peaks that were not present than on the peaks that were present; hence the samples with the fewer peaks were clustered together. By treating these samples as a sub-set of the data and performing the PCA on only those peaks present in the three different chromatographic traces (eighteen peaks) a better classification can be made. The two step process is, however, not ideal. It requires an increase in the amount of time spent interpreting the data and requires the operator to perform a subjective interpretation of the results of the initial PCA. The desired situation is a one step chemometric classification of the samples.

6.1.1.5 Classification of ECD raw data using K-means clustering

The raw data for the ECD analysis of the samples of HFC 134a from the known sources and the blank samples of air and helium contained a total of 43 cases defined by 46 variables. The values for the peak heights ranged from below LOD (0.015) to the largest peak (1014071 for Hoechst sample peak 26). From the previous cluster analysis performed using K-means clustering to investigate the FID chromatographic raw data (6.1.1.1) it is obvious that the large differences in peak heights in the ECD data will have a major effect on the cluster patterns formed.

A K-means cluster analysis was performed on the raw data set using nine cluster centers (the expected number of sample groups). This cluster analysis correctly classified both the UK and Atochem samples into distinct groups. However, because of the number of large peaks with relatively large peak heights in the Hoechst samples these lead to a bias in the clustering over the samples with smaller peaks. The Hoechst samples were split into five separate groups each containing one sample. All the other samples were placed into one large group. Further analysis of the raw data was therefore not performed as the separation of the samples was highly influenced by the range of peak heights.

6.1.1.6 Classification of ECD normalised data using K-means clustering

The data for the chromatographic analysis of the known samples of HFC 134a and blanks by ECD were normalised by the same method as described in 5.2.2. The normalised data set was subjected to K-means cluster analysis using 9 to 14 cluster centres to describe the data. The classification results are displayed in Table 6.5.

The classification of the samples using nine cluster centres shows that all the samples can be distinguished, apart from the Japan samples and the Helium blanks. The Hoechst samples were split into two distinct clusters. Using eleven clusters to describe the data allows all the samples except for one sample from Japan to be correctly classified according to their origin of production. Further increases in the number of clusters used to describe the data set does not improve the classification

pattern showing the consideration of the optimum number of PCs required to describe the data to be very important[8].

Table 6.5 *K- means classification of normalised ECD chromatographic data using nine to fourteen cluster centres. The first, second and third columns of numbers indicate the number of correctly, distinctly and mis classified samples using that number of cluster centres, respectively*

Sample origin	Number of clusters used to classify samples					
	9 clusters	10 clusters	11 clusters	12 clusters	13 clusters	14 clusters
UK	5	4 1	4 1	4 1	3 2	3 2
USA	5	5	5	5	5	5
Japan	5	5	4 1	4 1	4 1	4 1
Atochem	5	5	5	5	5	5
Hoechst	5	4 1	4 1	3 2	3 2	2 3
Showa Denka	5	5	5	5	5	5
Ashai glass	5	5	5	5	5	5
Air	4	4	4	4	4	4
Helium	4	4	4	4	4	4
Classification results						
Correctly classified	29	32	40	39	38	37
Distinctly classified	5	2	2	3	4	5
Mis-classified	9	9	1	1	1	1

A two step classification was applied to the ECD data in an attempt to classify the Japan sample into its correct group of origin. As in Section 6.1.1.4 the data representing the Japan samples and the Helium blanks were considered in isolation. The data was considered first as the normalised results produced using all the data and then as a data set normalised considering only the values for the Japan and Helium chromatographic traces.

Considering the initial results for the normalisation of the data and applying K-means clustering using two cluster centres the Japan sample previously misclassified is still grouped with the Helium samples. However, the distance measure from the cluster centre for the Helium samples is the highest in its group. Therefore applying a three cluster model leads to the misclassified Japan sample being placed in a group on its

own. The distances between these clusters is shown in Figure 6.4, where the distance between the main Japan cluster and the Helium cluster has been normalised to one.

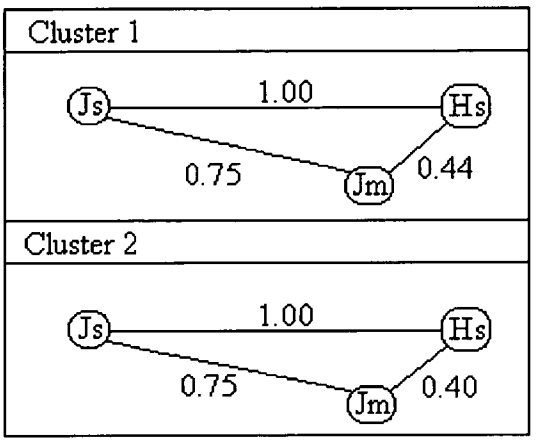


Figure 6.4 *Distance between cluster centres for K-means clustering of normalised data set for Japan samples and helium blanks where Js is the group of four Japan samples, Hs is the group of helium blanks and Jm is the misclassified Japan sample. Cluster 1 shows the separation of the samples using the normalised data from the data set of all the ECD results. Cluster 2 shows the separation of the samples using data normalised only for the values from Japan and Helium samples.*

The raw data for the samples representing Japan and Helium were taken and normalised using only the highest values from this reduced data set. These normalised values were then subjected to a K-means cluster analysis using two and three cluster centres. The classification of the samples did not change when compared with the previous normalised data results (see Figure 6.4).

On examining the raw data (Appendix H) for the misclassified Japan sample compared with the other samples from Japan a general trend can be seen. The peak height values for the outlying Japan sample are consistently lower. The peaks seen in all the other samples from Japan are seen in the outlying sample but the heights are approximately one quarter the size. The reason for this lower level may be because of an improved purity of the sample or an error in the analysis of the sample reducing its concentration. The Japan sample is misclassified as a Helium blank for a number of reasons. The overall levels of impurities in the Japan samples are low. Hence these samples are clustered closely to the Helium blanks which have few peaks present. In the case of the misclassified sample the reduced level of the by-products present places the sample closer to the 'zero' value of 0.015 than to the values for the other Japan samples; hence the sample is misclassified. It is obvious, however, from

a simple inspection of the chromatograms for each ECD analysis that the sample is closely related to the Japan samples. Therefore there is a problem in using this method of classification on the data set when the numbers of descriptive peaks are both few and low in response.

6.1.1.7 Classification of ECD PC scores using K-means clustering

The principal component scores for the analysis of the correlation coefficient matrix for the ECD data for the samples of HFC 134a were clustered using K-means clustering. The ECD data set was clustered using various numbers of PCs to describe the data and with a range of cluster centres to find the optimum separation parameters.

The number of PCs which provided the best separation was investigated. The results of performing the K-means cluster analysis on three to eight PCs using nine cluster centres are shown in Table 6.6. Describing the data using 7 PCs provided the best separation of the data with only nine misclassified samples. Increasing the number of PCs did not improve the separation as these values only added random variation into the classification system.

Using seven PCs to describe the data the PC scores were subjected to a K-means cluster analysis using a range of cluster centres (nine-fourteen) to describe the data. The results for altering the number of cluster centres from nine to fourteen are displayed in Table 6.7. The K-means clustering which supplied the best separation was achieved using seven PCs defined by ten clusters which misclassified only two samples and distinctly classified only one. The two misclassified samples were from Japan and Ashai Glass. One of the Japan samples was again misclassified with the Helium samples because of the reasons previously discussed in Section 6.1.1.4.

Table 6.6 *K- means classification of PC scores for ECD chromatographic data ranging the number of PCs from 3 to 8.*

Sample origin	Number of clusters used to classify samples					
	3 PCs	4 PCs	5 PCs	6 PCs	7 PCs	8 PCs
UK	2 3	2 3	4 1	4 1	2 3	2 3
USA	5	5	5	5	5	5
Japan	5	5	5	4 1	5	5
Atochem	5	5	5	5	5	5
Hoechst	5	5	5	5	5	1 4
Showa Denka	5	5	5	5	5	5
Ashai glass	5	5	5	5	4 1	5
Air	4	4	4	4	4	4
Helium	4	4	4	4	4	4
Classification results						
Correctly classified	23	23	27	31	32	33
Distinctly classified	7	2	1	1	2	3
Mis-classified	13	18	15	11	9	12

Table 6.7 *K-means classification of the first seven PC scores for ECD chromatographic data.*

Sample origin	Number of clusters used to classify samples					
	9 clusters	10 clusters	11 clusters	12 clusters	13 clusters	14 clusters
UK	2 3	4 1	4 1	4 1	5	5
USA	5	5	5	5	5	5
Japan	5	4 1	4 1	4 1	4 1	4 1
Atochem	5	5	5	5	5	5
Hoechst	5	5	4 1	4 1	5	5
Showa Denka	5	5	5	4 1	4 1	5
Ashai glass	4 1	4 1	4 1	4 1	4 1	1
Air	4	4	4	4	4	4
Helium	4	4	4	4	4	4
Classification results						
Correctly classified	32	40	39	38	35	31
Distinctly classified	2	1	2	3	6	10
Mis-classified	9	2	2	2	2	2

The Ashai Glass sample is misclassified into the Japan cluster. However, the Ashai Glass sample can be separated from the Japan samples by increasing the number of

cluster centres used to define the samples to fourteen. However, this is generally detrimental to the overall classification of samples. Table 6.8 shows the values for the distance measures between the cluster centres used to describe the data. The table represents the distance matrix between the nine cluster centres which is symmetrical about its diagonal, the diagonal values being equal to zero. From the table it can be seen that the misclassified samples are in the clusters which are close together, five and six (Japan and Ashai Glass) and six and seven (Japan and Helium). Both of the misclassified samples are switched between these close groups. Therefore the actual distance by which these samples are misclassified is low. Applying a different cluster analysis technique may therefore allow correct separation of the samples. Clusters 1 (UK) and 6 (Japan) are positioned closely but contain no misclassified samples from each others groups suggesting these data points are close together in the multi-dimensional space which they are projected into. Clusters 1 and 2 are close together, as would be expected, as these represent samples from the same origin split into two subsets.

Table 6.8 *Distance between cluster centres (ten clusters) for K-means clustering of the first seven ECD PC scores. Where each cluster represents a of samples from a specific origin 1=UK, 2=UK, 3=Hoechst, 4=Atochem, 5=Ashai Glass, 6=Japan, 7=Helium, 8=Air, 9=Showa Denka, and 10=USA. Numbers in bold in the table indicate cluster containing a misclassified sample.*

Cluster	1	2	3	4	5	*6*	*7*	8	9
2	**2.50**								
3	3.54	5.45							
4	3.51	5.41	4.08						
5	3.66	5.69	4.04	4.00					
6	**2.08**	4.37	3.33	3.37	**2.06**				
7	3.30	5.49	3.91	3.89	4.13	**2.44**			
8	3.87	5.43	4.30	4.30	4.21	3.64	4.06		
9	3.78	5.16	4.05	4.05	4.04	3.13	3.88	4.27	
10	3.68	5.32	4.08	4.08	4.09	3.24	3.93	4.30	4.07

6.1.2 Classification of known samples using hierarchical clustering

The hierarchical clustering techniques reported here are based on an agglomerative scheme[1], that is starting from all the samples being separated and linking each sample in turn until the samples are classified as one large group (see Section 2.2.3).

The two parameters that were altered between different hierarchical methods were the way the cluster centres were calculated and the measurement of distance between the samples[9].

6.1.2.1 Classification of FID raw data using hierarchical clustering

The data set for the FID analysis of the known samples of HFC 134a was subjected to a hierarchical cluster analysis using average linkage between groups with Euclidean distance measures. The values for the distance measures between the samples are displayed as a dendrogram in Figure 6.5 where the case numbers represent UK (1-5), USA (6-10), Japan (11-15), Atochem (16-20), Hoechst (21-25), Showa Denka (26-30) and Ashai Glass (31-35). The dendrogram shows that the samples representing the USA have been placed in a distinct cluster and are the last group to be merged with the rest of the samples. Moving from right to left the dendrogram shows how the samples have been clustered together. The next group represents the samples from Hoechst (cases 21-25) followed by the samples from Atochem (cases 16-20). The UK samples are also placed into a distinct group (cases 1-5).

The Japan, Showa Denka and Ashai Glass samples have all been clustered together in a close group. The scale on the dendrogram does not allow the samples to be clearly distinguished. However, by looking at the values in the agglomeration table the precise separation of the samples can be determined.

The agglomeration schedule showed that cases 29 and 30 were linked in stage 1 followed by the linking of these two samples (now described as cluster 29) to case 27 in stage 2. Case 26 representing a Showa Denka sample is linked to case 31, an Ashai Glass sample, leading to a misclassification. This cluster (26 and 31) was then linked to the cluster representing cases 27, 28, 29 and 30 in stage 13. This agglomeration procedure links all the samples from Showa Denka into one group with the misclassification of one Ashai Glass sample.

The cases representing Japan were numbered 11-15 and were first clustered in stage 4 where cases 14 and 15 were linked together. Cases 11 and 13 were linked together in stage 6 and so form two groups within the Japan group. The cluster containing cases 14 and 15 was linked to case 12 in stage 16. However, before the two groups of Japan samples were linked into one group the method joined group 12 (12, 14 and 15) with group 32 (32, 33, 34 and 35) containing Ashai Glass samples. This was followed by the linking of group 12 with group 26 (26, 27, 28, 29, 30 and 31) representing another misclassification. All these groups were then joined to form one large cluster.

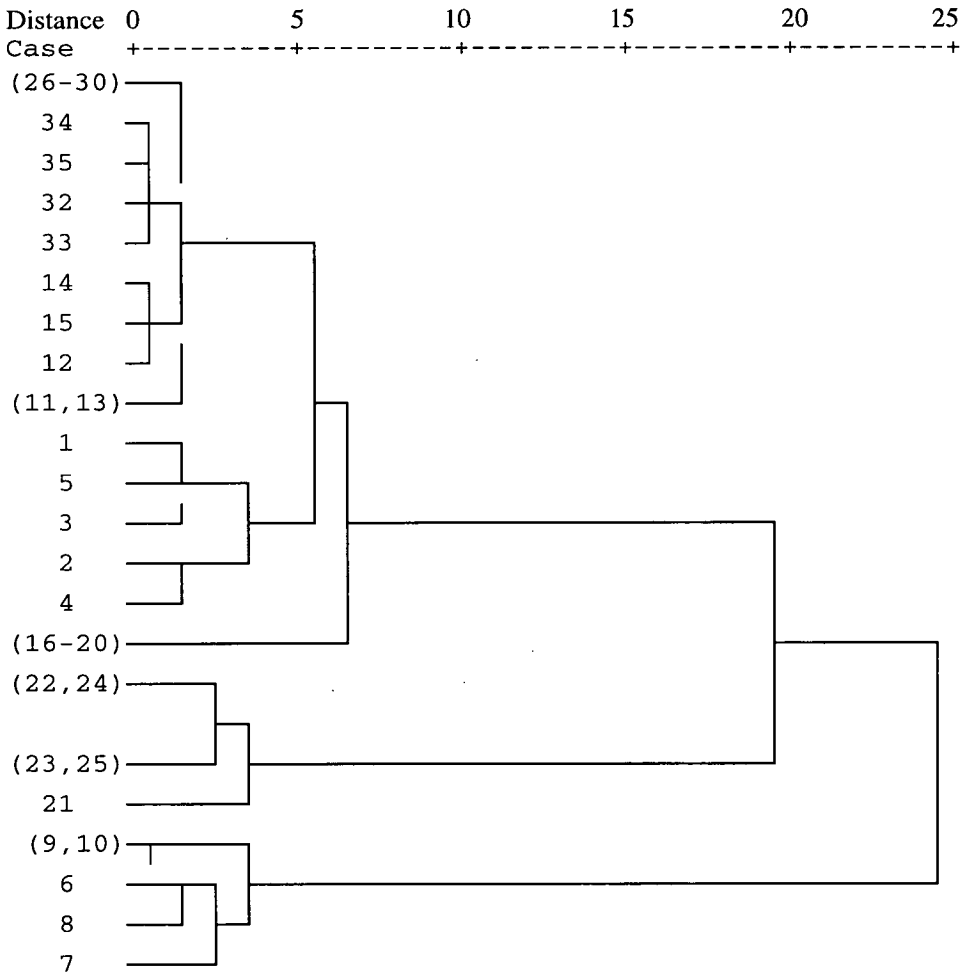


Figure 6.5 *Dendrogram of FID raw data using average between group linking with Euclidean distance measures.*

This classification procedure therefore misclassified one Ashai Glass sample into the Showa Denka cluster, misclassified three Japan samples to link with the Ashai Glass

samples and classified the remaining Japan samples as a group distinct from these groups.

The data were subjected to a second hierarchical clustering schedule using the same between groups average linking but using the squared Euclidean distance measure. The agglomeration schedule differed only slightly and only after the point when a number of samples had been defined by new cluster centres. The Japan samples were all clustered into one discrete group. Cases 26 (Showa Denka) and 31 (Ashai Glass) were misclassified into a group together before being linked to the other cases representing Showa Denka. This technique therefore led to the misclassification of only one sample (case 31, Ashai Glass). Further hierarchical clustering was not performed as better descriptions of the data were available using the normalised and PC analysed data of the FID chromatographic results.

6.1.2.2 Classification of FID normalised data using hierarchical clustering

The normalised data for the FID analysis of the known samples of HFC 134a were analysed using hierarchical clustering. A range of different distance measures were used to calculate the between samples and between cluster distances. Different calculation methods were also employed to define the cluster centres after joining samples together. In all combinations of distance measures and cluster definitions the samples were not fully separated. The samples for Japan, Showa Denka and Ashai Glass were misclassified, to varying degrees, by each method employed while all other samples were clearly separated.

6.1.2.3 Classification of FID PC scores using hierarchical clustering

The FID PC scores obtained from the correlation coefficients matrix were subjected to hierarchical clustering. As with the K-means cluster analysis the number of PCs used was varied in order to ascertain the optimum number of components required to produce clear separation of the samples (6.1.1.3).

A hierarchical cluster analysis was performed on the FID data described by the first five PCs using squared Euclidean distance measures with average between group linking. As with the raw and normalised FID data clear separation could be achieved for the UK, USA, Atochem and Hoechst samples. In addition the Showa Denka samples could be separated from the Japan and Ashai Glass samples. The dendrogram for the hierarchical clustering is displayed in Figure 6.6 where the case numbers represent UK (1-5), USA (6-10), Japan (11-15), Atochem (16-20), Hoechst

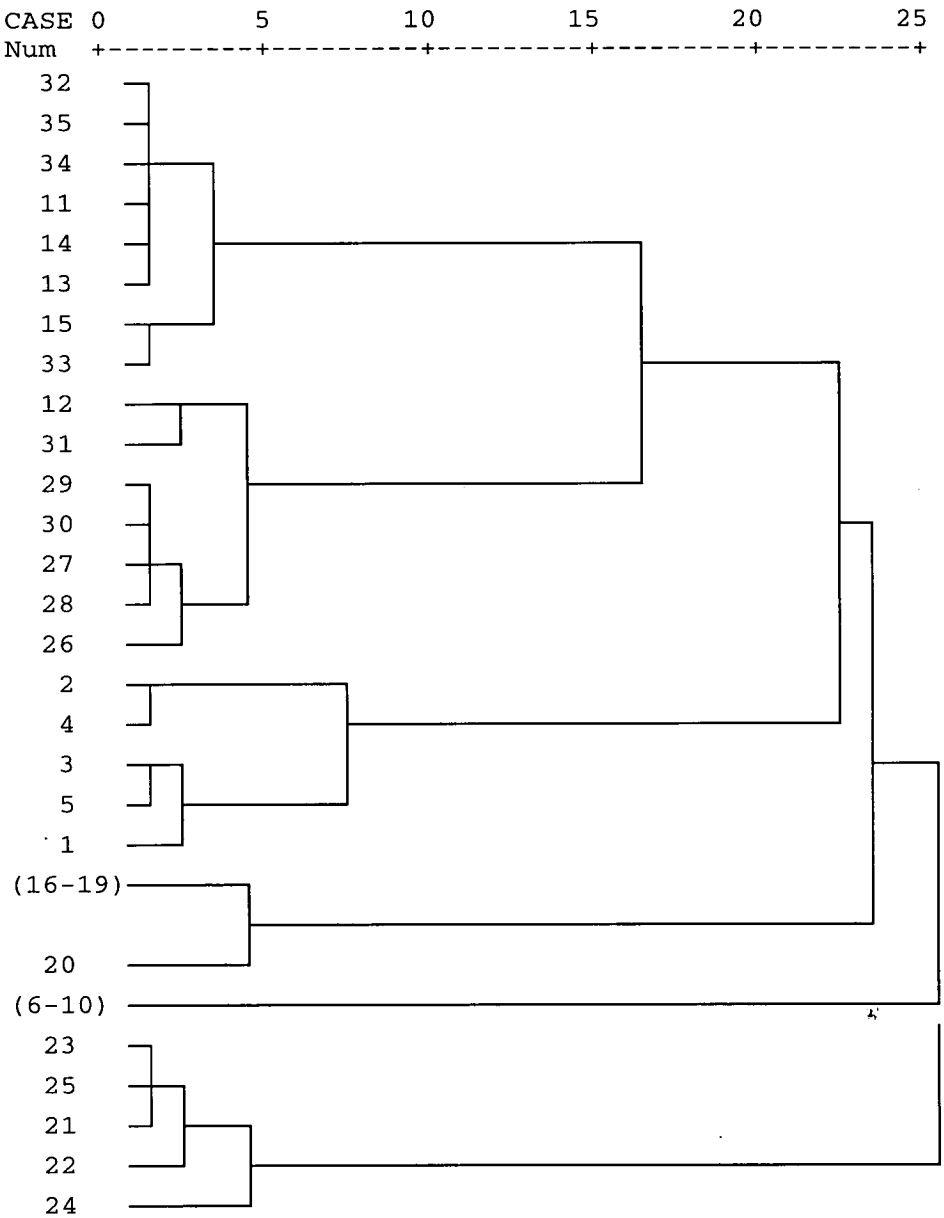


Figure 6.6 *Dendrogram of hierarchical clustering results for FID using the first five PC score data.*

(21-25), Showa Denka (26-30) and Ashai Glass (31-35). The Showa Denka samples are placed into a distinct group but this group is associated with two misclassified cases (12 and 31). The effect of this association depends on the scale at which the clusters are defined. The different distance level used to define a cluster dramatically alters the way in which the data is interpreted. Table 6.9 shows the effect of altering the distance at which the groups are defined on the cluster membership and therefore how well the samples are separated.

Table 6.9 *Effect of altering the distance limit (possibly between 0 to 25) in relation to cluster definition of samples. Case numbers represent* UK (1-5), USA (6-10), Japan (11-15), Atochem (16-20), Hoechst (21-25), Showa Denka (26-30) and Ashai Glass (31-35). *Misclassification based on definition in Section 6.1.1.3.*

Distance	Number of clusters	Cluster number	Cases in each cluster	Classification of samples		
				Correct	Distinct	Misclassified
25	1		All samples			35
20	5	1	11-15, 26-35			15
		2	1-5	5		
		3	16-20	5		
		4	6-10	5		
		5	21-25	5		
10	6	1	32-35, 11, 13-15			8
		2	26-30, 12, 31			7
		3	1-5	5		
		4	16-20	5		
		5	6-10	5		
		6	21-25	5		
3	10	1	32-35, 11, 13-15			8
		2	12, 31			2
		3	26-30	5		
		4	2, 4		2	
		5	1, 3, 5		3	
		6	16-19	4		
		7	20		1	
		8	6-10	5		
		9	21-23, 25	4		
		10	24		1	

At a distance of 25, the re-scaled distance at which point all the samples are agglomerated into one cluster, all the samples are indistinguishable from one another and therefore considered misclassified. At a distance of 20, the UK, USA, Atochem and Hoechst samples are placed in separate groups whereas the Japan, Showa Denka

and Ashai Glass samples are misclassified into one group. Reducing the distance to 10 has no effect on the already correctly classified samples but does separate the mixed samples into separate groups (group 2 contains predominantly Showa Denka samples). Reducing the distance to 3 allows the separation of the Showa Denka samples into a cluster separate from all other samples. However, at this distance sub-groups within the already separated samples of UK, Atochem and Hoechst appear, reducing the overall separation efficiency.

6.1.2.4 Classification of ECD raw data using hierarchical clustering

The data set for the ECD analysis of the known samples of HFC 134a was subjected to a hierarchical cluster analysis using average linkage between groups with Euclidean distance measures. Initial analysis of the data showed that the cases representing the Hoechst samples were highly different from all the other samples. The absolute difference led to a limited amount of information being displayed in the dendrogram. That is, the re-scaled distance between the groups of samples was dwarfed by the difference between the Hoechst samples and all other samples.

The hierarchical clustering was repeated without the Hoechst samples to allow the sub-structure within the samples analysed to be viewed. The dendrogram formed from the agglomeration of the samples is shown in Figure 6.7. The UK samples (1-5) are shown as a distinct group which has been separated from all the other samples. Samples representing USA (6-10), Atochem (16-20), Air (31-34) and Helium (35-38) have all been correctly placed into groups. The samples from Japan (11-15), Showa Denka (21-25) and Ashai Glass (26-30) were all misclassified into either a mixed group or placed into other clusters (*e.g.* case 11 into the Helium cluster).

The use of different distance measures and different functions to describe the new cluster centres did not lead to any significant improvement in the separation of the samples. As seen with the K-means clustering of the ECD data the very large differences in peak heights have a major effect on the cluster results. This was

highlighted by the fact that the cases which contain the largest peak height values were the easiest to separate.

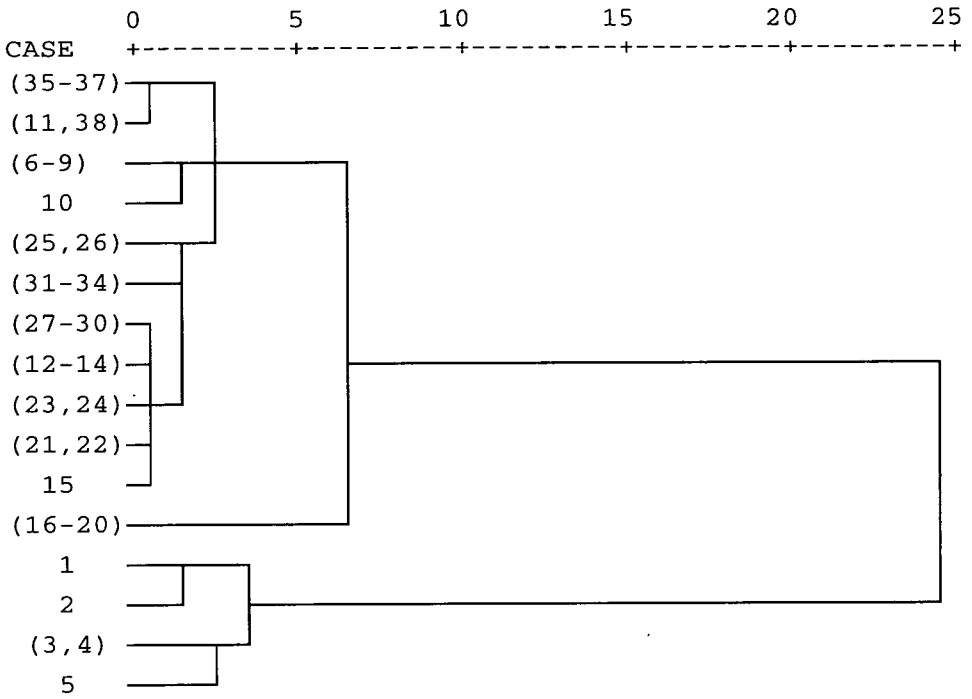


Figure 6.7 *Dendrogram for hierarchical clustering of ECD raw data using average between group linkage with Euclidean distance measures.*

6.1.2.5 Classification of ECD normalised data using hierarchical clustering

The normalised data for the analysis of the different samples of HFC 134a were subjected to a hierarchical cluster analysis using Euclidean distance measures with between group linkage. The agglomeration schedule is displayed as a dendrogram in Figure 6.8.

The first obvious improvement over the ECD raw data clustered using the hierarchical method is that all the samples can be displayed on one dendrogram and still retain a large amount of information in the sub-structure. The difference between clusters and within clusters can be viewed for all the samples analysed.

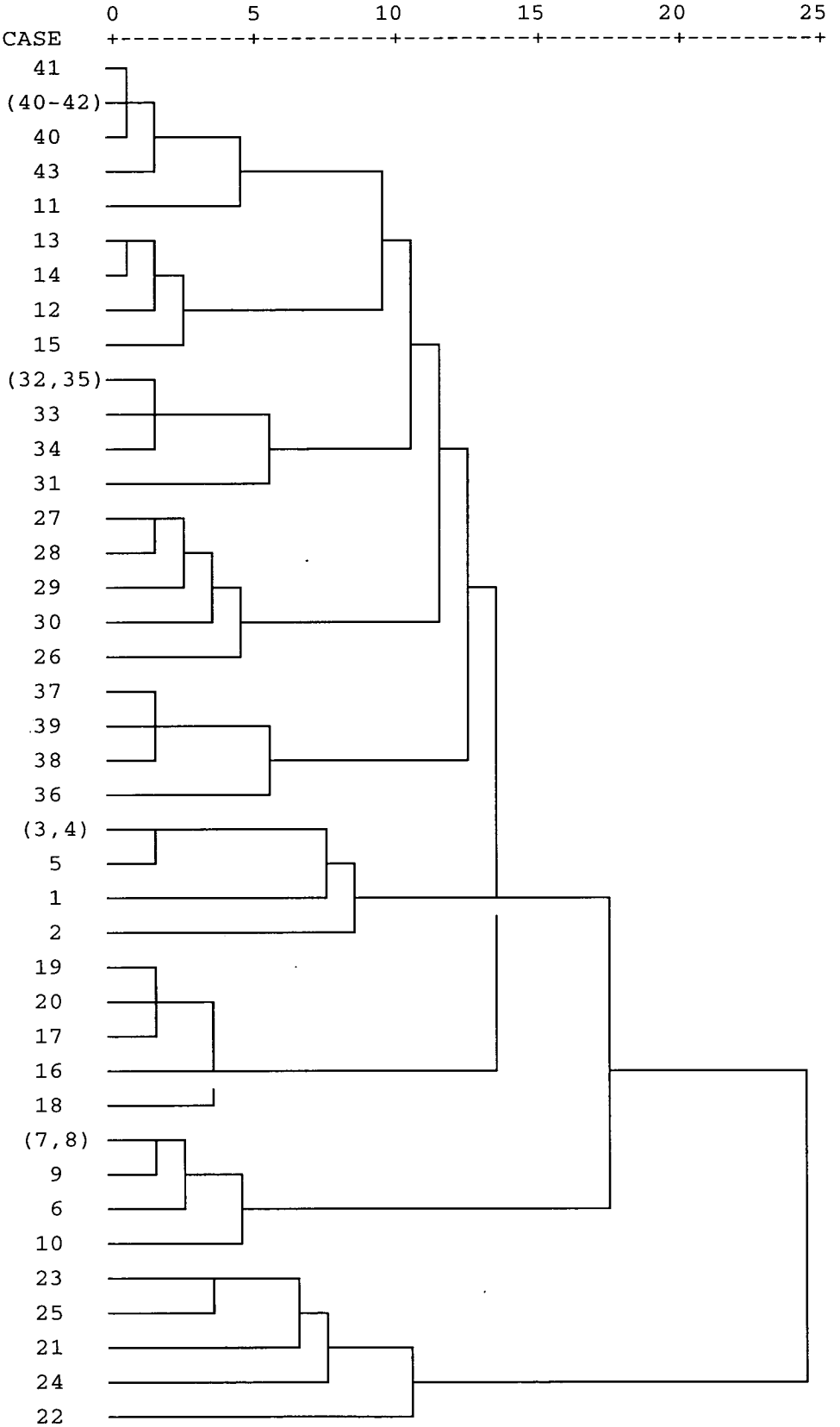


Figure 6.8 *Dendrogram for hierarchical clustering of ECD normalised data using Euclidean distance measure with average between cluster centres.*

The samples representing Hoechst (21-25) are clustered into a distinct separate group and are seen as the most different from all the other samples. This result matches that seen for the ECD raw data. The cluster pattern for the remaining samples shows that all the samples have been placed into distinct clusters on the basis of their origins where the cases represent 1-5 UK, 6-10 USA, 11-15 Japan, 16-20 Atochem, 26-30 Showa Denka, 31-35 Ashai Glass, 36-39 air and 40-44 Helium.

The use of the Euclidean distance measure to classify the different cases into distinct samples has not separated the samples completely. It is therefore necessary to use a different type of distance measure. The cosine measure is a pattern similarity measure for continuous data[10]. The calculation measures the cosine of the angle between two vectors of values. The cosine ranges from -1 to +1, with a value of 0 indicating orthogonal vectors *i.e.* unrelated. Figure 6.9 shows the dendrogram formed by an agglomerative hierarchical cluster analysis using cosine distance measures with average between group linkage.

The similarity measure between the different cases as measured by the cosine method ranges shows the most similar samples to have measurements of close to 1. That is the angle between the two vectors which are used to describe the samples is close to zero. Values approaching 0 are those defined by cases which have vectors approaching 90^{o} which are orthogonal and therefore unrelated.

The dendrogram shows that the samples have all been classified into distinct groups representing the different origins of the samples. Case 11 which has been consistently misclassified using other methods of similarity has been placed into the correct grouping. However, the nature of this similarity classification measure needs to be considered (see Section 2.2.1.1).

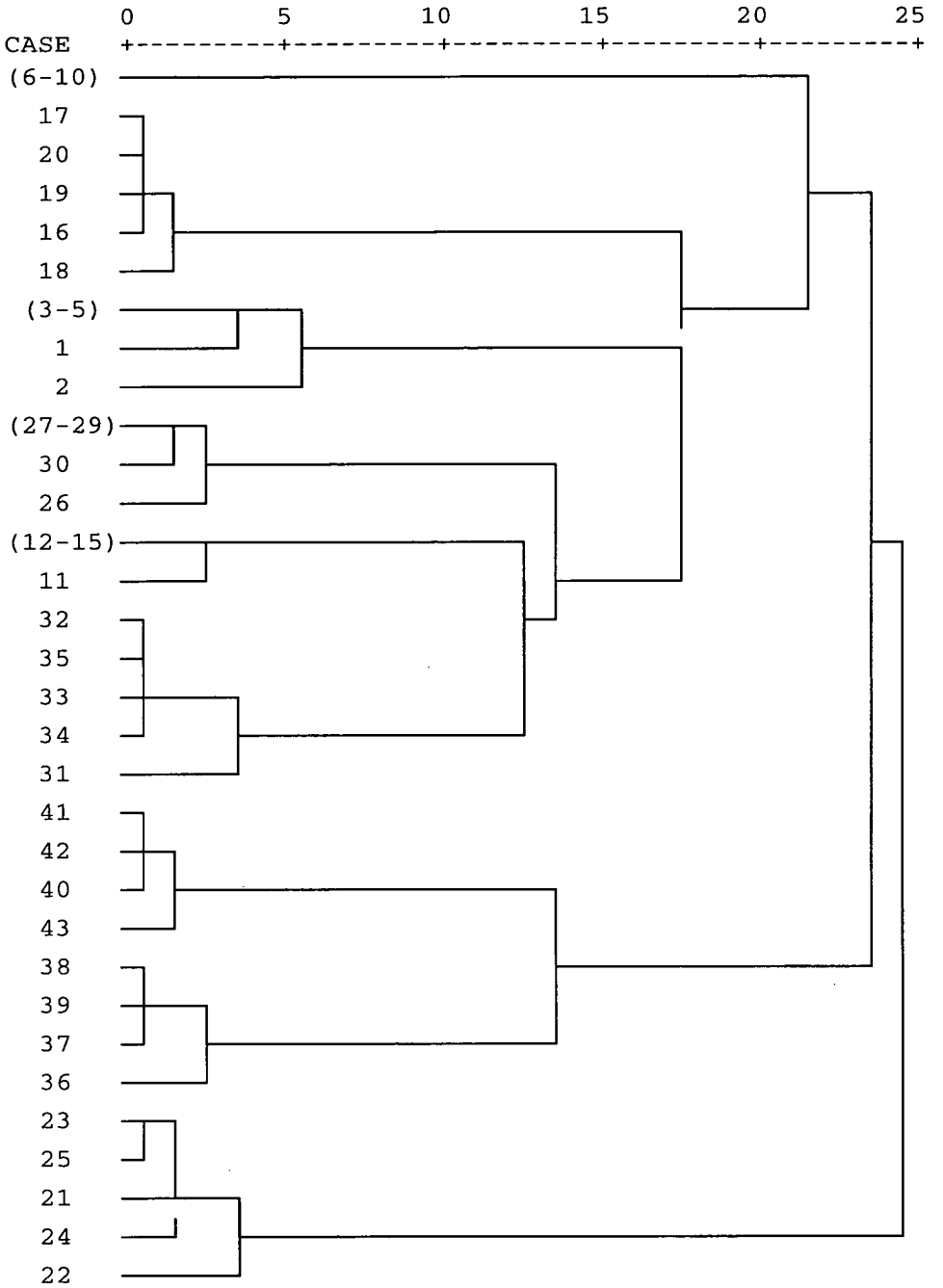


Figure 6.9 *Dendrogram of ECD normalised data using cosine measurements with average between group linkage.*

6.1.2.6 Classification of ECD PC scores using hierarchical clustering

The PC scores extracted from the correlation coefficients matrix of the ECD data were subjected to hierarchical clustering. The distance measures and cluster centre definitions were altered to obtain the best possible separation. The number of PCs was also altered to find the minimum number of variables needed to describe the data

so that all the samples could be correctly classified on the basis of their origins of production.

The results of the hierarchical clustering of the ECD data described by various numbers of PCs using the Euclidean distance measure with average between group linking are shown in Table 6.10. The table shows that by altering the number of PCs used to describe the data the efficiency of the separation can be changed. The optimum number of PCs used to describe the data was shown to be eight (describing 90.7% of the original variance of the ECD data). Using fewer PCs did not provide enough information on the original variance within the samples to allow a complete separation of the samples. Using more than eight variables led to a deterioration in the description of the data as these PCs describe only the residual random variation within the samples leading to the separation of samples from within the same production origin groups. This method of classification using eight PCs correctly classified all but one sample from Japan (Figure 6.10).

Further classification analyses were performed using different measurement methods and definitions for the cluster centres of the ECD samples to try to correctly classify the Japan sample (number 11). Further analyses were unable to correctly classify the Japan sample into its correct group (place of origin). However, a number of the techniques did improve the clustering of the samples using different cluster centre definitions. Figure 6.11 shows the dendrogram for the hierarchical clustering of the ECD data using eight PCs, Euclidean distance measures with Ward's method (section 2.2.2) for the calculation of the cluster centres. This method leads to the formation of tighter clusters (compare Figure 6.10 and 6.11) although does not improve the actual cluster membership *i.e.* the Japan sample 11 is still misclassified into the Helium cluster.

Table 6.10 *Hierarchical classification of PC scores for ECD chromatographic data ranging the number of PCs from 5 to 10.*

Sample origin	Number of clusters used to classify samples					
	5 PCs	6 PCs	7 PCs	8 PCs	9 PCs	10 PCs
UK	5	5	5	5	5	4 1
USA	5	5	5	5	5	5
Japan	4 1	4 1	4 1	4 1	4 1	4 1
Atochem	5	5	5	5	5	5
Hoechst	5	5	5	5	4 1	4 1
Showa Denka	4 1	4 1	5	5	5	5
Ashai glass	4 1	4 1	4 1	5	5	5
Air	4	4	4	4	4	4
Helium	4	4	4	4	4	4
Classification results						
Correctly classified	40	40	41	42	41	40
Distinctly classified					1	2
Mis-classified	3	3	2	1	1	1

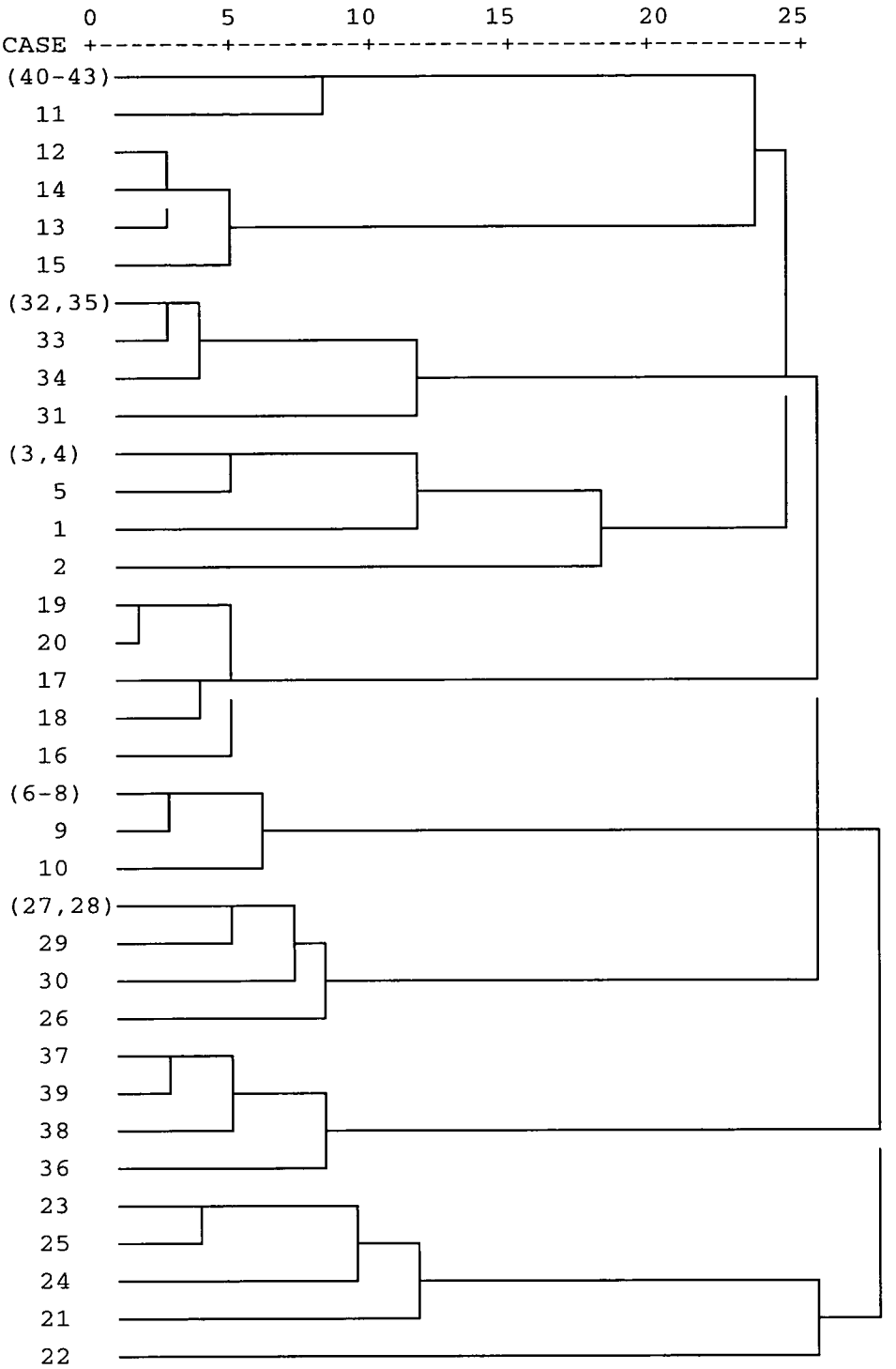


Figure 6.10 *Dendrogram of ECD data described by eight PCs using Euclidean distance measures with average between group linkage.*

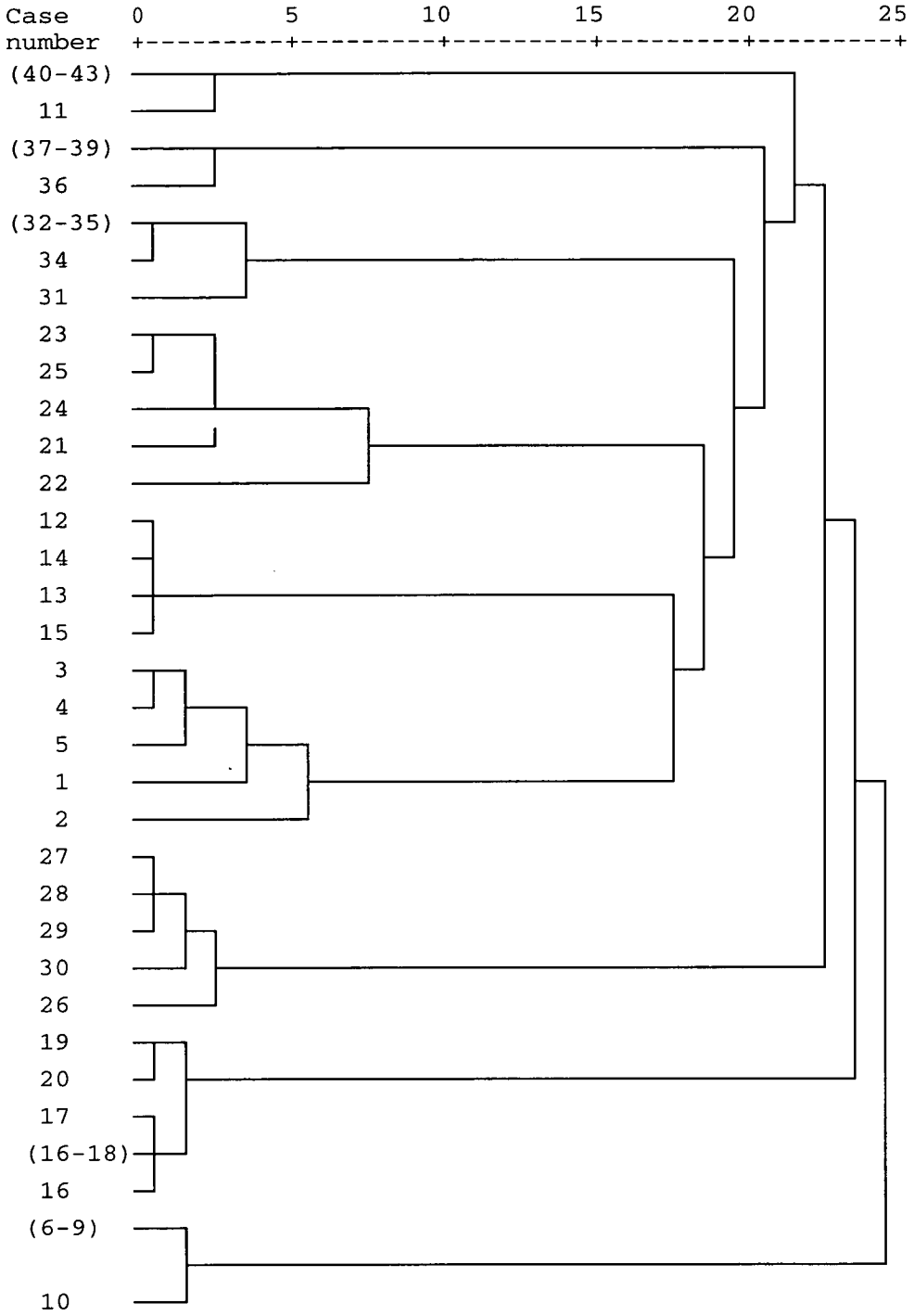


Figure 6.11 *Dendrogram of ECD data described by eight PCs using Euclidean distance measures with Ward's method for cluster centres.*

6.2 Choice of clustering technique

The following sections summarise the findings for the different classification techniques applied to the different data sets employed to describe the samples of HFC 134a from known sources. The relative pros and cons of each method of

classification as applied to each data set are considered and the methods which are shown to provide the most information with respect to the separation of the samples are then utilised for the classification of unknown samples.

The cluster techniques employed in the previous sections have provided extensive information about the nature of the samples. The techniques separated the samples into distinct clusters on the basis of their origins of production to greater or lesser extents. However, in most cases the underlying trend was to correctly identify the samples from the UK, USA, Atochem, Hoechst and in the case of the ECD data both Air and Helium. The samples which were consistently misclassified were the Showa Denka and especially the Japan and Ashai Glass samples. Inspection of the raw data for these samples shows that they are similar in the fact that they are characterised by a smaller number of peaks and at lower levels than the other samples. In the case of the ECD analysis the fact that peaks are present separates them from the sample blanks of Air and Helium.

The effect of having fewer peaks to describe the samples means that these samples (Japan, Showa Denka and Ashai Glass) were being classified more on the basis of the peaks they did not contain rather than the peaks that they did contain. As seen in a number of classification methods these samples are misclassified into sample groups distinct from the other samples but not from one another. However, further analysis of these samples as a sub-set of the data often led to these samples from Japan, Showa Denka and Ashai Glass being separated.

The data sets were treated in two general formats because of the sub-set of samples which contained limited amounts of information. The first was as seven distinct groups which were defined by their origin of production and new and unknown samples were classified into these groups. The second method was to treat the data as five distinct groups initially, where one group consisted of the samples from Japan, Showa Denka and Ashai Glass. Any samples placed into this group were

subjected to a further chemometric analysis where only the cases in the sub-set group were considered.

6.2.1 Cluster analysis of raw data

In both data sets from the FID and ECD no chemometric analysis technique correctly distinguished the samples based on the raw data. In all cases samples were either misclassified into different production origin groups, or the groups of different samples were not distinguished from each other. It was, however, possible to distinguish between groups when dealing with direct comparisons. Comparing the Hoechst samples with the Showa Denka samples leads to two distinct groups. Comparing the Japan sample with Ashai Glass samples leads to two distinct groups using a number of different techniques when applied to the raw data.

Difficulties in classification arose when comparing the larger data set which consists of a number of different groups. The variation within the absolute values of some of the larger peaks seen in some of the samples is greater than the relative variation of peak heights between two sample groups. For example, in the ECD raw data Peak 13, the USA samples have an average peak height of 13 and the Japan samples have an average of 2. Comparing these two peaks shows an obvious difference; however, considering the Hoechst sample with a peak height of 520 where the difference in distances between the highest and the lowest is 30 the USA and Japan samples now appear close together and were not clustered as different samples.

The larger peaks present in a number of the samples dominate the separation techniques employed with the result that instead of being a true multivariate analysis of some forty variables the method becomes a classification process based on a reduced number of the larger peaks.

The raw data were therefore not considered particularly useful for the separation of the samples. None of the chemometric techniques will be applied to the raw data in order to distinguish/identify the unknown samples in Chapter seven. However, the

raw data are still very important when investigating the reasons behind why the transformed data sets have been classified and therefore why samples are similar or dissimilar.

6.2.2 Cluster analysis of normalised data

The results of applying the chemometric analysis techniques to the normalised data sets allowed the majority of samples to be classified into their groups on the basis of the location of their production. The FID and ECD chromatographic data both provided enough information to allow such classification of the samples. The normalisation of the data removed the effect of the absolute magnitude of peaks present in the different samples. This meant that the samples were classified in a truly multi-variate way without weightings being applied to peaks on the basis of their size. Performing this data transformation does mean, however, that a certain amount of the variation and information is lost from data. Care must be taken in drawing conclusions from the transformed data and it is usually necessary to refer back to the raw data to get a clearer picture of why the samples have been classified. By doing this it is possible to draw conclusions regarding which peaks are important for the distinguishing of samples which will allow an evaluation as to whether these peaks are related to the production method of the sample.

6.2.3 Cluster analysis of PC scores

Application of the classification techniques to the principal component scores for the FID and ECD data allowed the samples to be separated into distinct groups on the basis of their origins of production.

The classification techniques highlighted the fact there is an optimum number of PCs which should be used to describe the data. Trying to classify the samples using too few PCs means not enough of the variance within the original data is described. No set value can be placed on the amount of variance needed to describe the data before the PCA is performed as this value is highly reliant on the amount of variation in the data. That is, not all the variation in the data is useful but it is impossible to predict

how much of the variance is required to perform the classification. Once the number of PCs used to describe the data is increased past the optimum the extra variance supplied is of no real use. As the number of PCs increases the variation these PCs describe relates to random variations both between and within sample groups. The PC scores calculated from these PCs only add random variation to the classification method which leads to a deterioration in the classification pattern. The analysis performed in this work utilized training data sets and test data sets in order to determine the optimum number of PCs to describe the data, although methods such as cross validation can be employed when examining data[11,12].

The ultimate goal of the chemometric part of this project was to be able to visually represent, in an easy to understand format, the different samples of HFC 134a into groups on the basis of the sites of production. A one step method of data reduction did not allow the separation of the samples using only the first three principal components. However, by performing a two step PCA on, first, all the data and then secondly on any sub-set of samples not clearly separated by the first PCA, the samples could all be classified and this classification viewed on two separate graphs displaying the PC scores. These two step PCA classification techniques are used extensively in the following chapter to allow the identification of new and unknown samples of HFC 134a.

References

[1] Myors, R., Wells, R.J., Skopec, S.V., Crisp, P., Iavetz, R., Skopec, Z., Ekangaki, A. and Robertson, J., Analytical Communications, **35**, N12, p403-410, (1998)

[2] Brown, J.S., Crop Sci., **31**, N4, p915-922, (1991)

[3] Zhang, Q. and Boyle, R.D., Pattern Recogn, **24**, V9, p835-848, (1991)

[4] Wold, S., Chemom. Intel. Lab Syst., **2**, p37-52, (1987)

[5] Kendall, M.G., A Dictionary of Statistical Terms, 4th ed., Longman Group, UK, p208, (1982)

[6] Vodovotz, Y., Arteaga, G.E. and Nakai, S., *Food Res. Int.,* **26**, p355-363, (1993)

[7] Frank, I., and Kowalski, B.R., Chemometrics, *Anal. Chem.*, **54**, p232R-243R, (1982)

[8] Malinowski, F. and Howery, D., *Factor Analysis in Chemistry*, Wiley, New York, (1980)

[9] Manly, B., *Multivariate Statistical Methods: A Primer,* Chapman and Hall, London, (1994)

[10] SPSS for Windows, Version 9.0, SPSS Inc., Woking, England

[11] Wold, S., Technometrics, **20**, p73-77, (1982)

[12] Eastment, H. and Krzanowski, W., Tchnometrics, **24**, p73-77, (1982)

Chapter 7

Application of chemometric classification techniques

7.0 Introduction to chapter

The following chapter details how the chemometric process of PCA followed by hierarchical clustering has been applied to HFC 134a in order to classify their origins of production[1]. The samples which were classified can be split into three broad classes.

The first class of samples were known samples of HFC 134a that were deliberately mixed in specific ratios and used previously for the identification of peaks (Section 4.5). These mixed samples should be classified into the areas between their constituent samples, that is, a 50:50 mix of UK and USA samples should be classified approximately the same distance from each known sample cluster centre.

The second class of samples was made up of samples of HFC 134a from the known origins of production which had not been used in the classification data set. These samples were used to ascertain if the methods of experimental analysis and chemometric analysis employed to classify the samples of the training set could be applied to untested samples. The correct classification of these new samples would show that the methods employed are applicable to samples outside the training set.

The third class of samples consisted of samples which were tested blind and samples which were obtained from outside sources (from the Chemical Engineering Department of The University of Edinburgh). The identities of the samples were unknown to the analyst but held on record. These results would therefore not be subject to operator interpretation in the classification process. The majority of the samples received from the Chemical Engineering Department were in gas cylinders without labels as to their origins of production. The analysis results of these samples could not be checked as the origins of production of most of these samples were unknown.

7.1 Classification of mixed samples of HFC 134a

The following sections detail the results of performing the chemometric classification techniques, developed in the previous chapters, on a series of samples prepared by mixing samples of HFC 134a from known origins in known ratios[2]. These samples had been previously used to aid in the assignment of by-product peak identities between the different chromatographic traces[3].

7.1.1 Classification of mixed samples analysed by GC-FID

The expected results from the classification of mixed samples of HFC 134a from known origins would be to place these samples between the groups of samples from which they were prepared. The distance from each group would be expected to be directly proportional to the mixture ratio. The most appropriate chemometric analysis technique which would provide the best format for the data to be viewed in was considered to be a PCA[4]. The PC scores which contained the most information relevant to the mixed samples could then be plotted to show the spread of data points.

A PCA was performed on a data matrix containing the FID chromatographic results for the known samples and the chromatographic results for the mixed samples. This matrix contained forty two cases described by thirty seven variables. The eigenvectors and eigenvalues were extracted from the correlation coefficients matrix to obtain the optimum results (see Section 5.3.1.2).

The PC scores for the first two components are plotted in Figure 7.1a. The total variance described by these two components is 68.4% compared with 70.3% when the known samples were considered in isolation. This difference reflects the variance added to the system by the new samples.

Figure 7.1a shows that the clusters of samples obtained when using only the known samples of HFC 134a are retained. That is, the first two principal components clearly separate the USA and Hoechst samples from all the others. An enlargement of the

scale around the area of the rest of the samples (Figure 7.1b) shows the level of separation achieved for these samples using the first two principal component scores.

The mixed sample for USA and Hoechst appears at a position approximately half way between the two clusters for the USA and Hoechst showing that this sample could be correctly assigned with no prior knowledge of its identity. Therefore, given no other knowledge about its identity apart from the fact that it is a mixture of two of the samples it would be a reasonable assumption to have identified the sample as a mixture of USA and Hoechst in a ratio of approximately 50:50. Comparing the simplicity of identifying the sample using these chemometric techniques with the situation of trying to manually identify chromatographic traces by eye illustrates the power of these techniques.

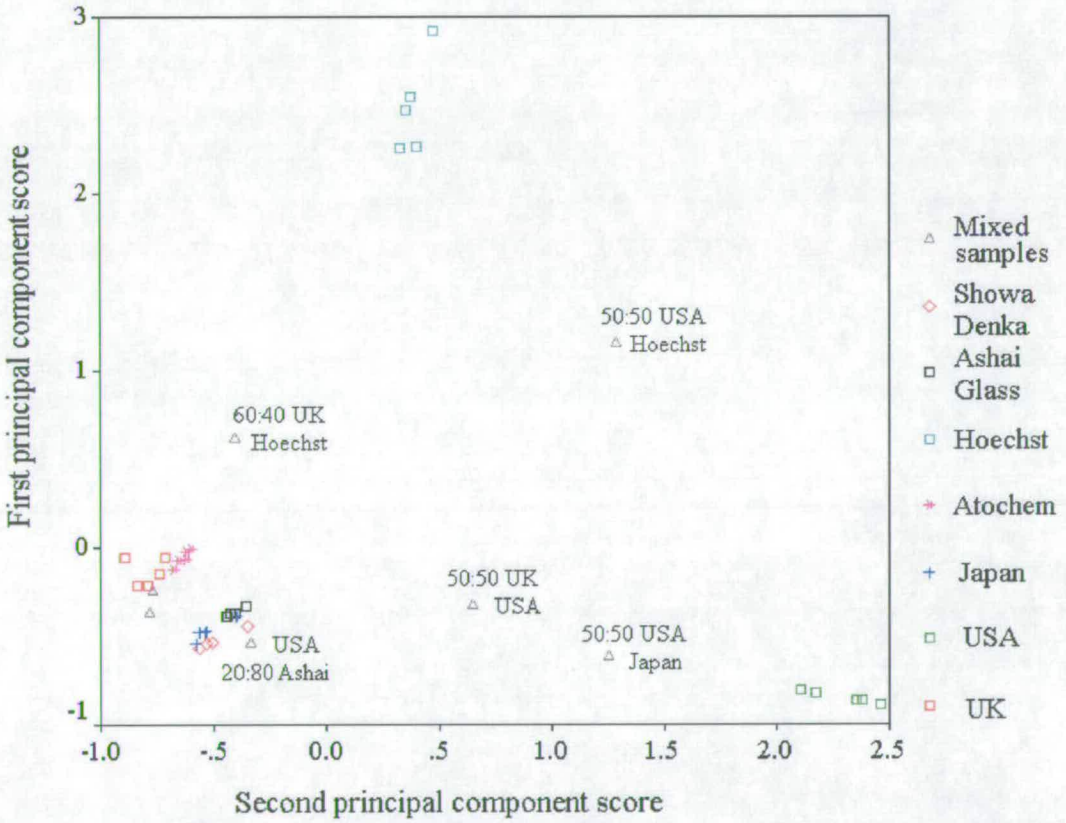


Figure 7.1a *First two PC scores for FID data of known origin and mixed samples of HFC 134a.*

The samples corresponding to the 50:50 mixes of USA:Japan and UK:USA both appear in approximately the correct positions on the chart. These samples, however,

need prior knowledge to be labelled. That is, it is possible to say that the mixture contains a percentage of the USA sample as it is positioned towards the USA cluster, but it is not possible from Figure 7.1a to correctly identify which sample it was mixed with. Similarly the UK:Hoechst and USA:Ashai weighted mixtures appear in the area expected on the graph but, without prior knowledge, both mixture identities cannot be ascertained.

The enlargement of the area corresponding to the rest of the samples shows in more detail the separation of the known samples and the positioning of the mixed samples. The variance needed to separate fully the samples shown in Figure 7.1b is described by further principal components and therefore the information contained within these components does not allow the correct identification of the mixed samples. It is therefore necessary to consider more of the PCs at the same time to allow the correct identification of the samples.

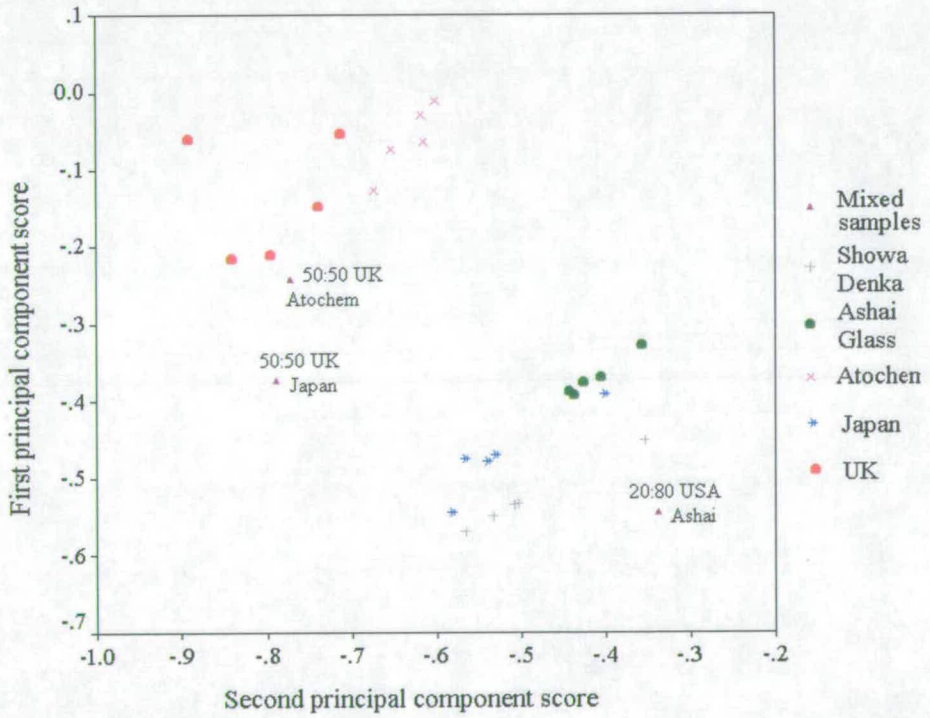


Figure 7.1b *First two PC scores for FID data of known origin and mixed samples of HFC 134a.*

The classification of mixtures of samples when they cannot be visualised is, however, beset with problems. Consider three sets of samples that are clearly

distinguished from one another in multi dimensional space (whether considered as raw, transformed or PC scores). The analysis of a 50:50 mix of two of these samples may place the mixture in a position in space between its two constituent samples. The position this mixture occupies can then be measured from each of the three sample groups. The problem arises if the third sample group (not part of the mixture) lies in close proximity to one of the mixture samples. The measurement of the mixture of the samples could then appear to be closer to the third group than to one of the actual mixture constituents giving the incorrect impression that the third group makes up part of the mixture. An example of this effect is seen in Figure 7.1a where the 50:50 mix of UK and USA would be measured as closer to the Showa Denka samples than to the UK samples.

Consideration of smaller numbers of sample groups in isolation can allow a clearer picture to be formed when classifying the samples. Figure 7.2 shows the results of a PCA using only the FID data describing the UK, USA and Atochem and 50:50 mixes of these samples. The figure shows the plot of the first principal component score against the second principal component score. The three groups of known samples are clearly distinguished with the 50:50 mix samples placed approximately between the two groups.

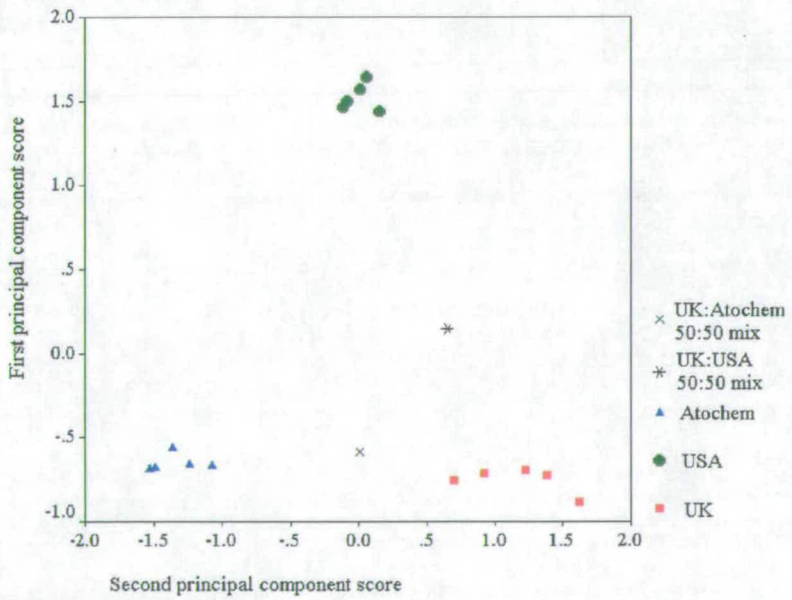


Figure 7.2 *Plot of first two PC scores for UK, USA and Atochem samples considered alone with two 50:50 mixes of the samples.*

When dealing with mixed samples, the problem may arise of being unable to easily distinguish between a mixture of two samples and a distinct group belonging to a new class of samples. One way to investigate such a problem would be to make up a series of mixtures of the different samples and then analyse these to show where the mixtures should appear. If an unknown sample lies on this curve between the two sample groups the sample could be a mixture of them. However, there is the other possibility that even though an unknown sample has a very similar position in multi-dimensional space as a blend of two samples it may still belong to a distinct origin of production which by coincidence has a similar by-product profile to that of a blend of samples.

7.1.2 Classification of mixed samples analysed by GC-ECD

The chromatographic results for the ECD analysis of the mixed samples of HFC 134a were investigated using chemometric techniques. PCA followed by the selective viewing of a number of the PCs allowed the visualisation of the samples on a series of graphs. Figure 7.3 shows the component scores for the first two components for the known samples and mixed samples of HFC 134a. The PCA analysis was performed using the correlation coefficients matrix which was calculated from the data matrix of forty eight cases defined by forty six variables.

The figure shows that the general separation of the samples of known origins is the same as that seen when analysing the data in the absence of the mixed samples. The positioning of the mixed samples on the figure is in the general areas that would be expected. For example, the 50:50 mixed sample of USA and Showa Denka appears approximately in the centre between the groups representing the constituent parts. The mixed samples for the USA:Japan and USA:Ashai Glass are positioned closer to the Japan and Ashai Glass samples as the mixtures were weighted (30:70 and 40:60, respectively). Similarly the mixed sample of UK:Hoechst was weighted 80:20 and therefore appears in Figure 7.3 much closer to the UK samples than to the Hoechst samples. The mixed sample of UK:Atochem (50:50) does not appear directly between the samples as the PCs used in this diagram do not contain enough of the

variance in respect to these samples. Viewing further PCs which highlight the differences between the UK and Atochem samples more clearly leads to a better positioning of the mixed sample.

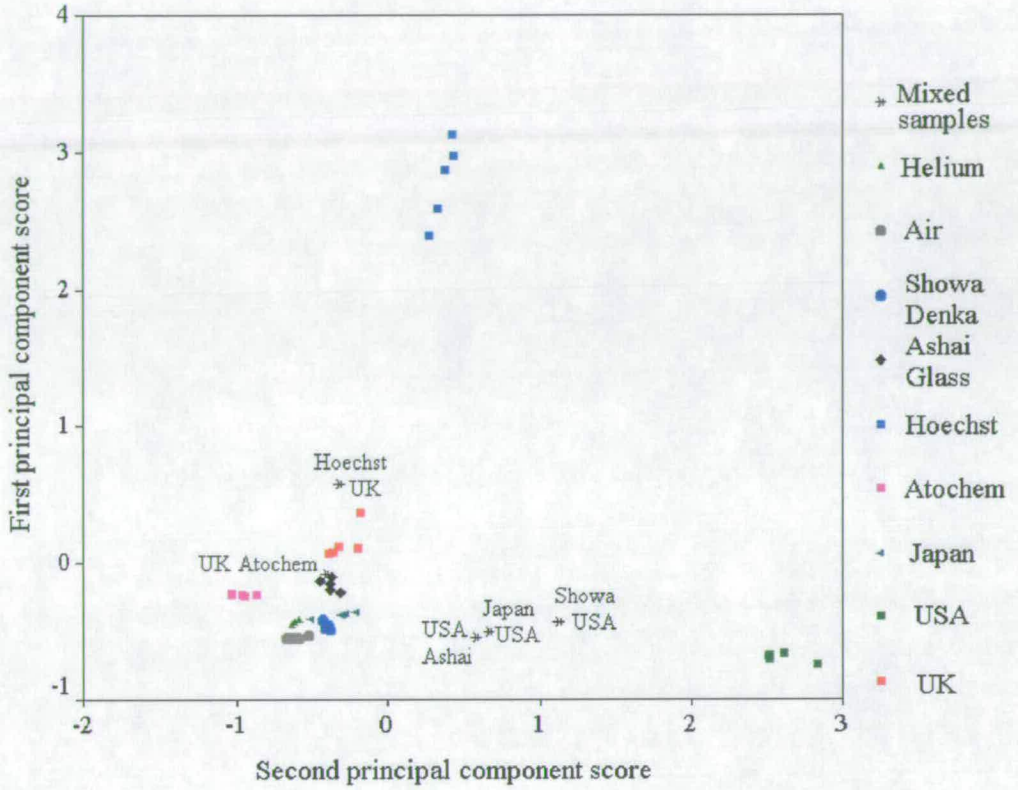


Figure 7.3 *PC scores for first two PC extracted from the ECD data of known samples and mixed known samples.*

Previous sections have shown that trying to describe all the samples with only a few PCs does not always provide a clear picture of the differences/similarities between the samples. By targeting a specific set of samples and performing the PCA on these samples in isolation, clearer descriptions of the samples can be made. One problem apparent in Figure 7.3 is how to distinguish between the mixed samples for USA:Japan, USA:Showa Denka and USA:Ashai Glass. Each of these mixed samples appears in the expected area (between the USA cluster and the general cluster containing Japan, Showa Denka and Ashai glass) but this area does not allow the samples to be distinguished. These data for the USA, Japan, Showa Denka, Ashai Glass and mixtures were taken and subjected to a separate PCA. The first three PC scores are displayed in Figure 7.4.

Figure 7.4 is a three dimensional plot of the component scores where the axes have been rotated to obtain a clearer view of the data. The PCA has separated the samples into their respective locations of production. The mixed samples all appear in a region of the graph as would be expected. However, the mixed samples do not lie on a straight line between each of the samples as would possibly be expected.

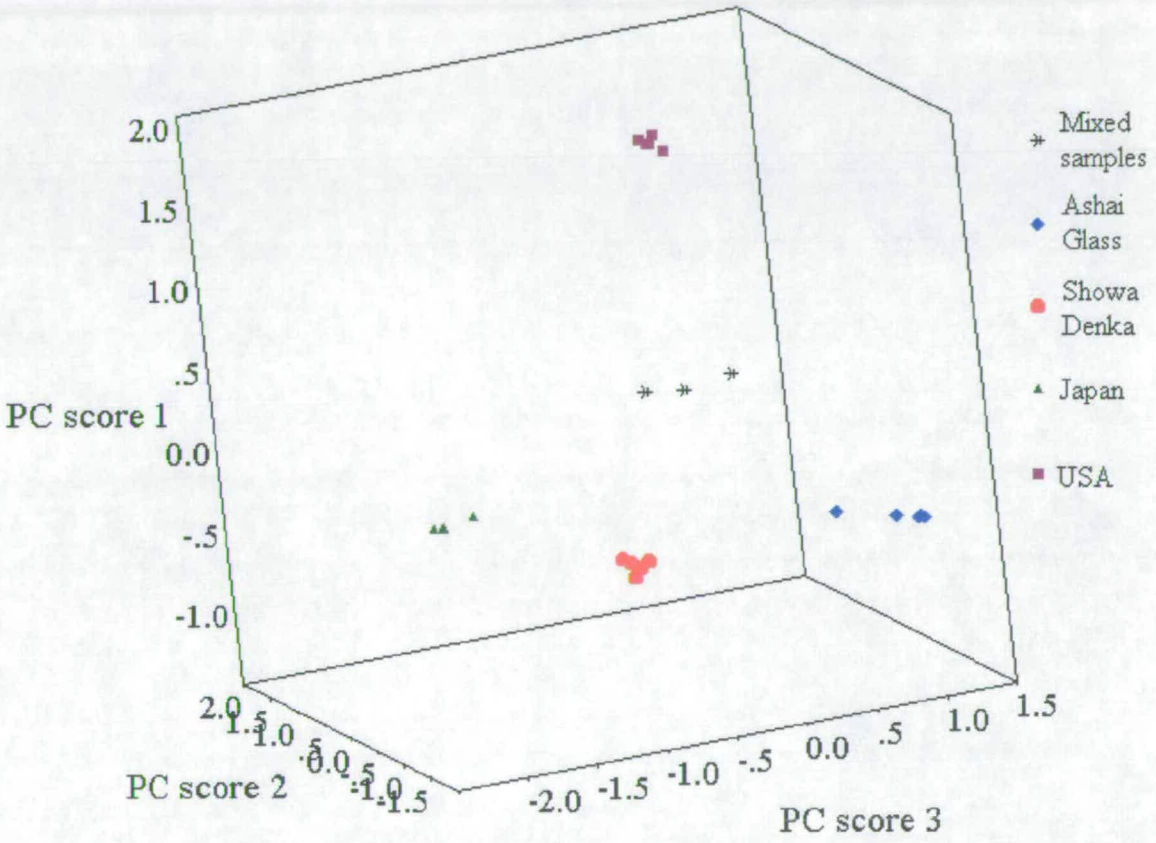


Figure 7.4 *Three dimensional plot of first three PC scores calculated from ECD data matrix for USA, Japan, Showa Denka, Ashai Glass and mixed samples in isolation.*

This is due to the variance described by each PC being uncorrelated (orthogonal axes, Section 2.1.7.2). For example, a mixture containing 100% Japan would appear in the Japan cluster; increasing the concentration of USA in the mixture will move the sample towards the USA cluster. However, the progress in each dimension is not necessarily linked; the reduction or increase in a peak will have more of an effect on the component score with which it is linked. Therefore from the limited information in Figure 7.4 it appears that the ability to distinguish the mixed samples contents with respect to Japan, Showa Denka and Ashai Glass disappears before the sample reaches

the USA cluster. This example again displays how a two step method to resolve the clustering of samples can be applied to provide the information required for correct classification of samples.

7.2 Classification of known origin test set of HFC 134a samples

The problem with the development and optimisation of classification protocols is that in many cases a method is formed which can correctly separate the test samples but this method does not work when applied to samples outside the test sample groups. In order to check the methods developed in this work, a series of different samples were obtained from the various production sites. These samples, when analysed, should be classified into the correct groups as defined by the training samples on the basis of origin of production.

Eleven samples of HFC 134a were obtained; two samples from ICI UK, two from ICI USA, one from ICI Japan, two from Atochem, two from Hoechst, one from Showa Denka and one from Ashai Glass. These samples were analysed by GC-FID and GC-ECD to obtain chromatographic profiles to be used in the classification of the samples. Component identities were made on the basis of retention time values (both absolute and relative to HFC 134a), relative detector response and mass spectral identification of the major components.

7.2.1 Classification of test set samples from known origin analysed by GC-FID

The data set which described the analysis of the known training and test samples of HFC 134a was entered into an SPSS spreadsheet. This data matrix consisted of forty six cases defined by peak heights for thirty seven variables. These data were subjected to a PCA where the eigenvalues and eigenvectors were calculated from the correlation coefficients matrix. The component scores for the first ten PCs were calculated for each case. A plot of the first two component scores (describing 71% of the original variance) for the different samples is shown in Figure 7.5a. The distribution of the points representing the different known samples is the same as that seen in Section 5.3.1.2. The samples produced by UK, USA, Atochem and Hoechst

are all clearly separated into distinct groups. The samples representing Japan, Showa Denka and Ashai Glass are not clearly separated when looking at the first two principal components (Figure 7.5b enlarged scale).

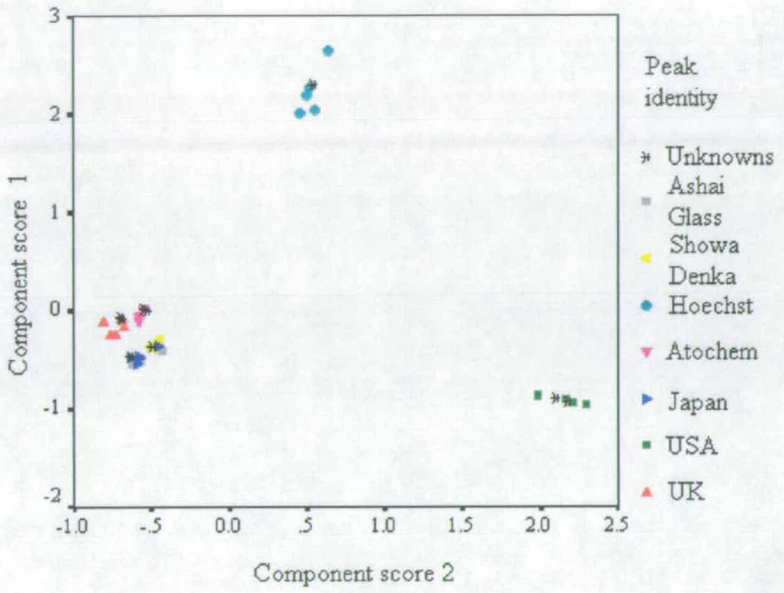


Figure 7.5a *First two PC scores plotted for FID chromatographic data representing the known and unknown test set samples of HFC 134a*

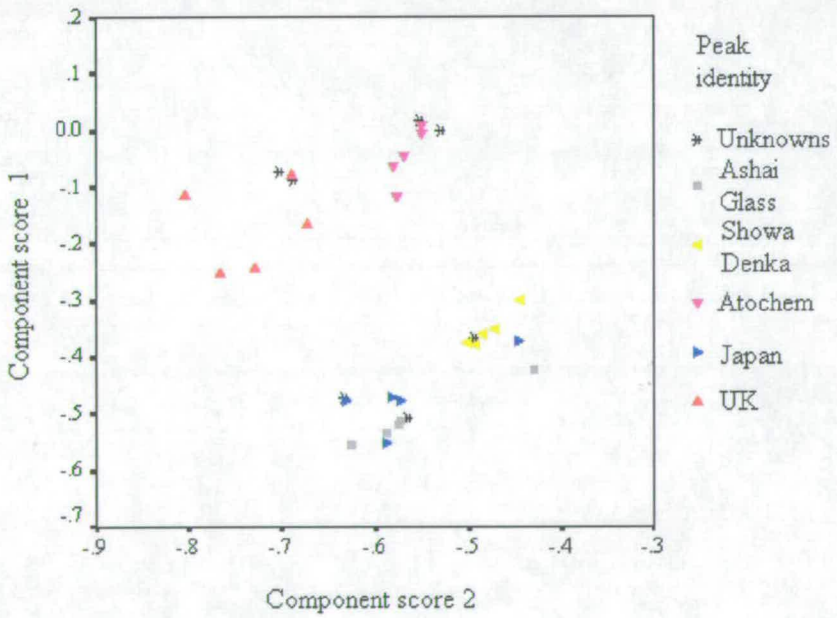


Figure 7.5b *First two PC scores plotted for FID chromatographic data representing the known and unknown test set samples of HFC 134a (enlarged scale).*

The test samples from the UK, USA, Atochem and Hoechst were all placed in close proximity to the clusters of the original samples. This shows that these samples have all been correctly identified as to their place of production.

To correctly classify the unresolved test samples into their origins of production it is necessary to perform an extra step. This can either involve using a classification technique which considers more of the component scores or by performing a second PCA on the unresolved cases. As has been shown in Chapter 6 the classification techniques applied to the FID data did not produce clear separation for the known samples analysed. Therefore a second PCA was performed on the data corresponding to the Japan, Showa Denka, Ashai Glass and unresolved test samples.

The data set corresponding to the Japan, Showa Denka, Ashai Glass and unresolved samples consisted of eighteen cases defined by seventeen variables. This reduced number of variables reflects the fact that these samples contained a low number of peaks, as detected by FID, in comparison with the samples from the other origins of production. The PC scores for the first two PCs extracted from the correlation coefficients matrix are displayed in Figure 7.6. The figure clearly shows that the samples from the different origins of production are separated into distinct areas. The cases representing the test samples have all be correctly associated with their place of origin.

The analysis of the known and test samples using a two step PCA has shown that cases can be correctly identified as to their place of origin. The method of chemometric analysis employed has been shown to be applicable to the identification of unknown samples from known origins. Section 7.3 will further test the methods to see if the same classification methods are applicable when looking at samples from potentially different sources by GC-FID and chemometrics.

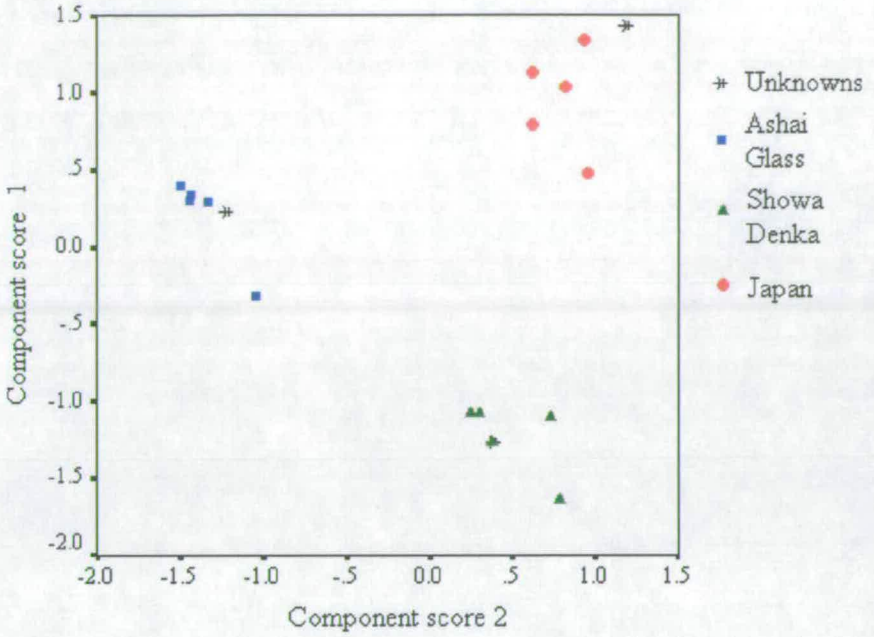


Figure 7.6 *PC scores plotted for FID chromatographic data of Japan, Showa Denka, Ashai Glass and unresolved test samples of HFC 134a.*

7.2.2 Classification of test set samples from known origin analysed by GC-ECD

The data for the analyses of the test samples were entered into an SPSS spreadsheet to form a data matrix of fifty four cases defined by forty six variables (peaks). These data were subjected to a PCA where the eigenvalues and eigenvectors were extracted from the correlation coefficients matrix. The component scores for the first six components were calculated for each case. PC score 1 is plotted against PC score 2 in Figure 7.7. 51% of the original variance is described by the first two PCs which separated most of the samples into their correct origins of production and correctly identified some of the test samples. For example, the test samples representing the production origins of ICI USA, Atochem and Hoechst have all been placed into the correct clusters which in turn are clearly distinct (Figure 7.7) from all other samples. The other known samples and test samples, however, are not as well separated. To correctly cluster all the samples, more of the original variance within the data needs to be considered.

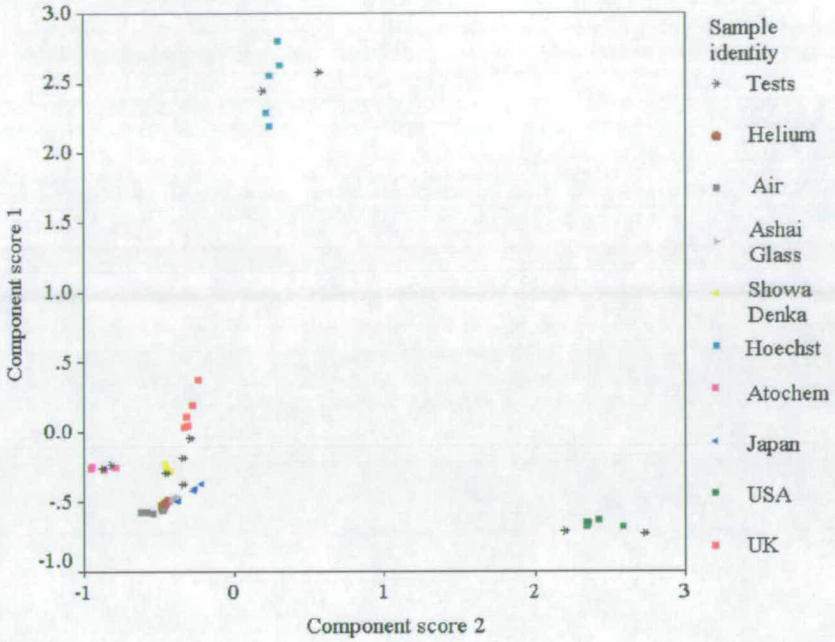


Figure 7.7 *PC1 and PC 2 scores for known and test data analysed by GC-ECD*

Section 6.1.2.6 showed that a hierarchical cluster analysis using approximately six PCs (80% of original variance in this matrix) to define the data allowed the full separation of the known samples into their origins of production. The agglomeration schedule for a hierarchical cluster analysis of the ECD data using Euclidean distance measures with Ward's method for cluster centre definitions is displayed in Appendix 7.1. The dendrogram representing this schedule is shown in Figure 7.8.

The training cases (samples of known origin) were labelled as follows: UK (1-5), USA (6-10), Japan (11-15), Atochem (16-20), Hoechst (21-25), Showa Denka (26-30), Ashai Glass (31-35), Air (36-39) and Helium (40-43). The new samples were labelled UK (44 and 45), USA (46 and 47), Japan (48), Atochem (49 and 50), Hoechst (51 and 52), Showa Denka (53), Ashai Glass (54). The first point to note is that the cases representing the samples from known origin, which were originally used to find the optimum method for clustering the data, have all been placed into distinct clusters on the basis of their origins of production. The exception is the Japan Sample 11 which was originally shown not to be correctly classified (see section 6.2 for explanation).

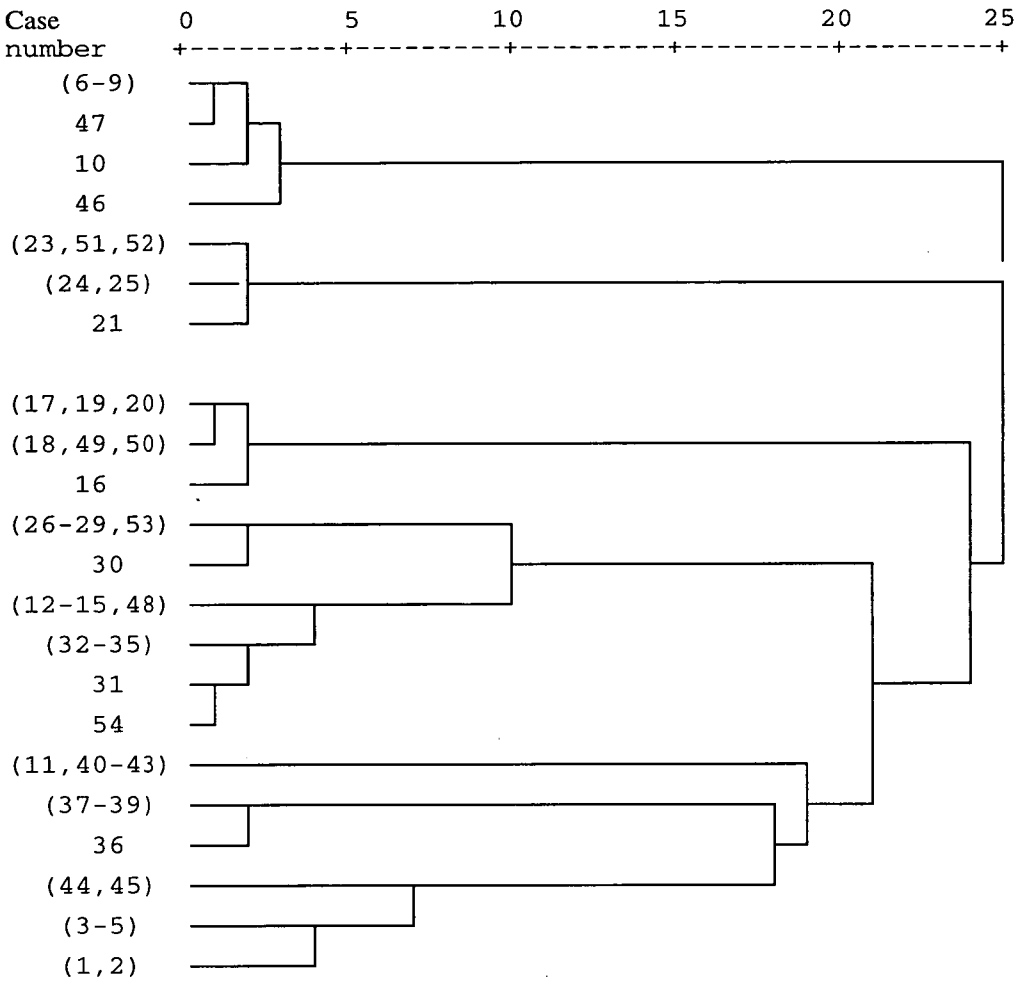


Figure 7.8 *Dendrogram of linkages for first six PC scores for known and test samples analysed by GC-ECD*

The test cases were all placed into the expected clusters of origin of production. Each of the test samples analysed was shown to be similar to the original training data therefore validating the method for a larger data set. The dendrogram does show, however, that in the case of the test UK samples (44 and 45) these displayed a level of variance which placed them into a sub-set of the UK origin cluster. If this signifies a real difference within this origin of production, further analyses of samples from this source will develop the difference to a more significant level.

7.3 Classification of samples of HFC 134a from unknown origin

The following two sections detail the results of applying the chemometric classification techniques to a set of samples of unknown origin (to the analyst)

analysed by both GC-FID and GC-ECD. These samples were obtained from ICI C&P Runcorn or from the Chemical Engineering Department of The University of Edinburgh. The samples received from ICI C & P were analysed blind in that the identities were withheld until a classification was made on the basis of the analysis of the samples. The samples from the Chemical Engineering Department were obtained without any identification to their origin of production and therefore the success of classification of these samples could not be confirmed.

7.3.1 Classification of samples from unknown origin analysed by GC-FID

The thirty five samples which represent the seven origins of production studied for HFC 134a were used to further classify samples from unknown origin. In total, fifty six analyses of unknown samples were performed using GC-FID. The peak assignment was made on the basis of retention time and relative retention time data. A number of the samples were analysed by GC-MS (see Section 4.4) and this information was used to align these peaks in the chromatographic trace. MS was not performed on all the samples because of a lack of instrument access time. However, this was not considered as a problem as one of the final applications of any techniques developed in this work would involve GC analysis without MS. The correct identification of samples without the added information supplied by MS would show that the techniques developed did not rely on MS assignment of by-product identity

The analysis of the unknown samples showed the presence of a peak not previously seen in any of the samples used in the training set. This peak was seen in the HQ samples (a sub set of the unknown samples obtained from the Chemical Engineering department) and was identified by MS as HFC 32 (difluoromethane). The peak chromatographed between peaks 22 and 23 and was therefore labelled Peak 22a to avoid altering the number labels of peaks with longer chromatographic elution times.

The unknown samples were split into groups for analysis on the basis of their place and time of acquisition. The samples were each chromatographically profiled using

the standard conditions by GC-FID. The chromatographic traces were transferred into an SPSS spreadsheet to form a data matrix containing the known and unknown data for HFC 134a samples. A final data matrix containing the chromatographic profiles of all the samples was prepared and considered after initial identification of the unknown samples was made. To initially consider the new data representing the unknown samples in one single chemometric analysis would have proved difficult and confusing.

7.3.1.1 Classification of unknown sample set Q1

The sample set Q1 represented a series of samples obtained from ICI C&P at the beginning of this work. These samples were received in a variety of gas cylinders and had no specific identification as to their origin of production. The sample set was made up of fourteen samples which were described by thirty eight variables. The peak height values were entered into the SPSS data matrix containing the results for the analysis of the training samples of known origin. The final data matrix contained forty nine cases defined by thirty eight variables (Peak 22a was not included as none of the cases had a value for this component).

The data were subjected to a PCA using the correlation coefficients matrix. The PC scores were calculated and used to classify the samples using a hierarchical cluster analysis using Euclidean distance measures and Ward's method to calculate the cluster centres. The results for the cluster analysis using the first six principal components (which describe 91% of the original variation in the data) are displayed in a dendrogram, Figure 7.9.

The cases representing the new samples have been split into three groups. Cases 46-49 have been placed into a distinct cluster which has been most closely associated with Cases 1-5 which represent the known samples of UK origin. It can therefore be concluded that although these samples are different from the UK samples they show a similarity to these samples. Cases 36-40 have been placed in a distinct cluster which is most closely associated with the cases representing the USA (6-10). The

final cluster of unknown cases (41-45) has been placed into a group containing all the samples representing Showa Denka (26-30) and one sample each from Japan (12) and Ashai Glass (31).

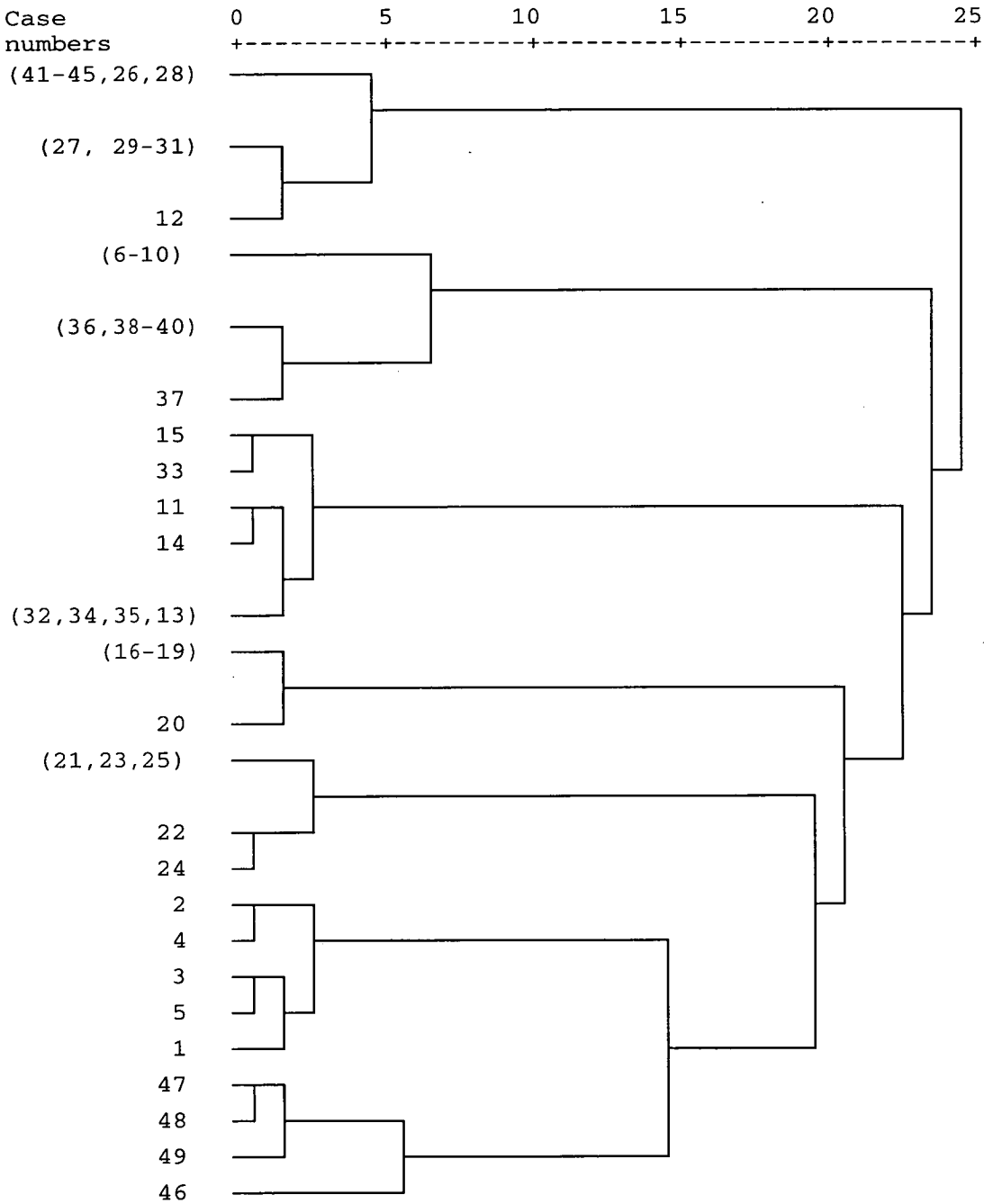


Figure 7.9 *Dendrogram of the hierarchical classification of PC scores calculated from data for known samples and unknown samples in group Q1 analysed by GC-FID.*

The samples which were not correctly classified using the first PCA were separated from the other samples and subjected to a second PCA in isolation. The PCA was

performed on twenty cases representing Japan, Showa Denka, Ashai Glass and the sub-population of the unknown samples (Q1). These samples were described by a reduced number of variables (seventeen). The first two principal component scores (86% of the original variation) were extracted and plotted in Figure 7.10. The figure clearly shows that the unknown samples originally placed into a group representing the Showa Denka samples (in PCA 1) are again placed in the sample group representing Showa Denka samples. The unknown samples display a high level of similarity and can therefore be considered as being from the same origin of production.

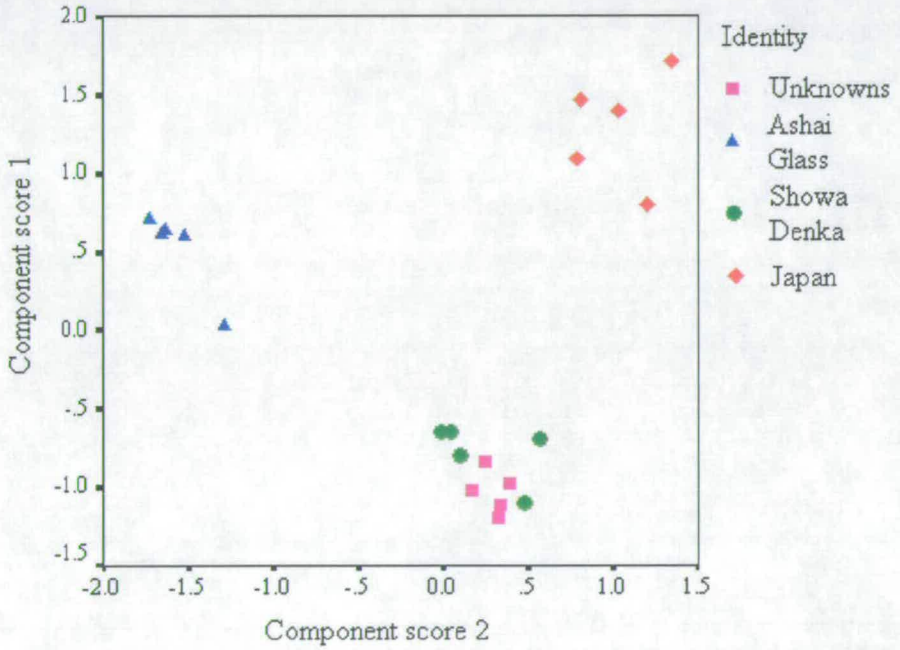


Figure 7.10 *PC scores for Japan, Showa Denka, Ashai glass and Unknown samples Q1 analysed by GC-FID.*

These results show that the unknown samples labelled Q1 contain five samples produced by Showa Denka, five samples which are closely associated with samples from the USA and four samples whose most similar samples are those corresponding to UK-produced HFC 134a.

7.3.1.2 Classification of unknown sample set Q2

The sample set Q2 represented a series of samples obtained from ICI C&P during the second year of this work. As with the previous set of samples they were received in a

variety of different gas cylinders. This set of unknowns consisted of sixteen samples described by thirty eight variables. The chromatographic profiles were aligned on the basis of retention time, RRT and mass spectral identification where available.

The data were entered into the data matrix containing the peak information for the samples of known origin. The samples representing Q1 were not included in the PCA at this stage. The data matrix was subjected to a PCA where the eigenvalues and eigenvectors were calculated from the correlation coefficients matrix. The PC scores for the first ten components were calculated. A hierarchical cluster analysis was performed on the first six PC scores (90% of the original variation) using Euclidean distance measures with Ward's method to calculate the cluster centres. The results for this analysis are displayed in the Figure 7.11 as an agglomerative dendrogram of linkages between the cases.

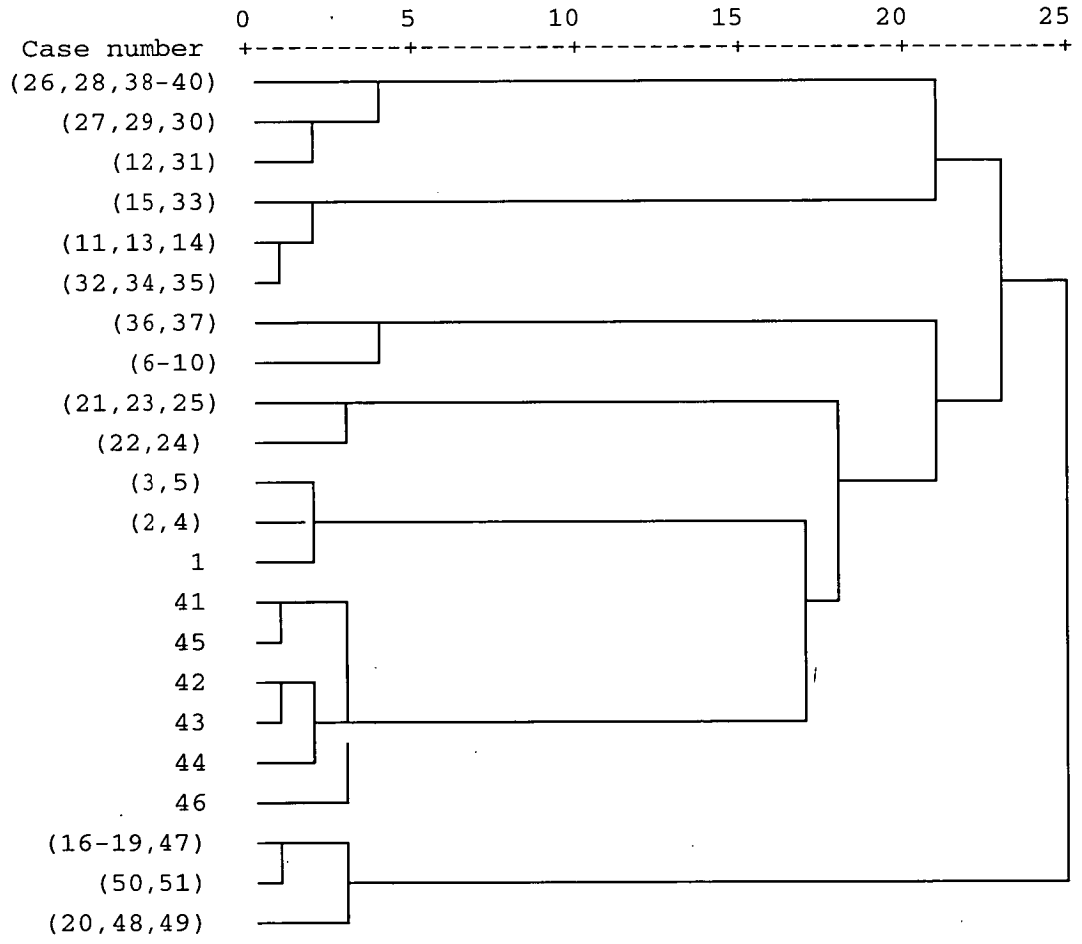


Figure 7.11 *Dendrogram of the hierarchical classification of PC scores calculated from data for known samples and unknown samples in group Q2 analysed by GC-FID.*

The classification of the samples shows that the known samples corresponding to Japan, Showa Denka and Ashai Glass have not all been correctly clustered. The pattern of misclassification matches that seen in the cluster analysis of unknown samples Q1 (section 7.3.1.1). However, a group of the unknown samples (labelled cases 38-40) have all been closely associated with the cluster containing mainly the Showa Denka samples.

A group of unknown samples Q2 (cases 47-51) have been clustered together with the samples representing Atochem (cases 16-20). The two unknown samples labelled cases 36 and 37 have been closely associated with the samples from the USA (cases 6-10). The unknown samples labelled 41-46 have been placed into a distinct cluster which is most closely associated with the samples from the UK. The samples which have not been classified into the correct clusters on the basis of their origins of production were subjected to a second PCA in isolation from those samples correctly classified.

The data matrix of the unresolved samples consisted of eighteen samples described by seventeen variables. The PC scores were calculated for the first two PC (74% of the original variation) from the correlation coefficients matrix and are plotted in Figure 7.12. The figure clearly shows that the three remaining unknown samples have been placed into the same sample cluster as those produced by Showa Denka.

The unknown samples labelled Q2 have been shown to contain three samples produced by Showa Denka, five samples produced by Atochem, two samples closely associated with USA-produced HFC 134a and six samples most similar to samples produced by ICI in the UK.

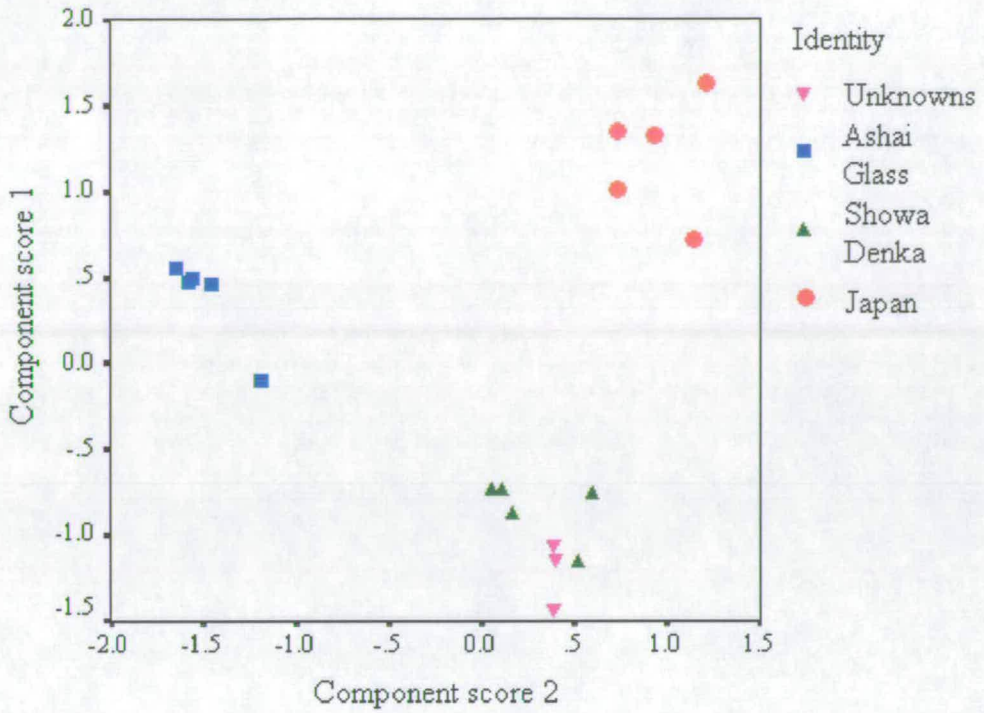


Figure 7.12 *PC scores for Japan, Showa Denka, Ashai glass and Unknown samples Q2 analysed by GC-FID.*

7.3.1.3 Classification of unknown sample set Q3

The sample set Q3 consisted of ten samples obtained in the third year of this work from ICI C&P. The GC-FID chromatographic profiles were determined for each sample and entered into the data matrix containing the results for the analyses of the training set of samples of HFC 134a. The data matrix consisted of forty five samples described by thirty eight variables. The eigenvectors and eigenvalues were extracted from the correlation coefficients matrix which was calculated from the raw data. The first ten PC scores were calculated for each of the samples with the first six PCs (which accounted for 92% of the original variation in the data) being used to define the data. A cluster analysis was performed on the first six PC scores for the samples analysed using a hierarchical method with Euclidean distance measures and Ward's method for cluster centre calculations. The results for this analysis are displayed in Figure 7.13 in the form of a dendrogram.

The dendrogram in Figure 7.13 shows the classification pattern of the training and unknown samples labelled Q3. The cases representing Japan (11-15) and Ashai

Glass (31-35) have not been clearly separated, with samples both mixed together and misclassified into other clusters. This does not affect the outcome of the cluster analysis, however, as the unknown samples have all been assigned to clusters defined by one origin of production.

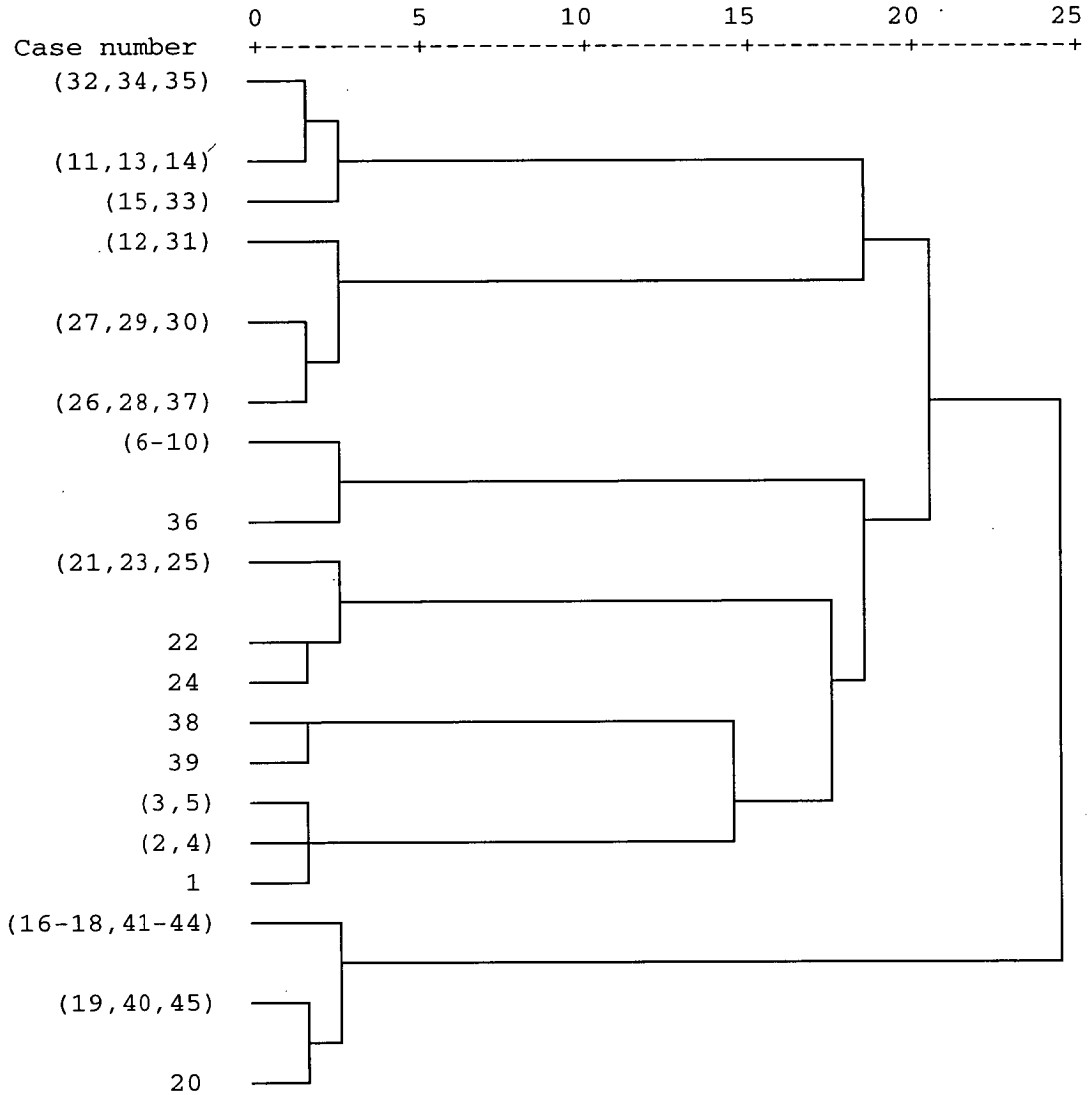


Figure 7.13 *Dendrogram of the hierarchical classification of PC scores calculated from data for known samples and unknown samples in group Q3 analysed by GC-FID.*

Case 36 (an unknown sample) was closely associated with the samples from the USA (6-10). Case 37 (an unknown sample) has been clustered with the samples from Showa Denka (26-30). Cases 40-45 (unknown samples) have been clustered with samples from Atochem (cases 16-20). The cases 38 and 39 (unknowns) have been

placed into a distinct group which is most closely associated with samples from ICI UK (cases 1-5).

7.3.1.4 Classification of unknown sample set CE

The sample set CE corresponded to those samples of HFC 134a obtained from the Chemical Engineering Department of The University of Edinburgh (courtesy of Dr Colin Pritchard). These samples were not labelled with an origin of production and in the main represented samples of HFC 134a left over from experiments relating to the refrigeration properties of HFC 134a. The success of the classification of these samples could not be assessed as the true identities of the samples were not known. However, the analysis and subsequent classification of the samples presented a 'real life' application of the techniques developed in this work when looking at unknown samples.

The data set consisted of seventeen unknown samples and thirty five samples of known origins. These samples were defined by thirty nine variables as three of the unknown samples from the CE set contained a peak not previously seen in any other sample (labelled 22a). The data matrix was subjected to a PCA and hierarchical cluster analysis on the first six PC scores using the same method as detailed in sections 7.3.1.1-3. The clustering results are displayed in Figure 7.14 as a dendrogram.

The unknown CE cases 50-52 have been clustered into the same group as the samples from Atochem (16-20). These three unknown samples have been clustered with cases 16-18 before cases 19 and 20 so they can be considered as definately identified as Atochem samples. This is because these unknown samples are more similar to a number of samples for Atochem than two of the training Atochem samples *i.e.* the differences between cases 50-52 are less than the intra variation between the cases known to originate from Atochem.

Cases 45 and 46 have been closely associated with the samples from the USA (6-10). The unknown cases 47 and 48 have been closely associated with the samples from

Showa Denka (26-30). Although the samples from Japan (11-15) and Ashai Glass (31-35) have not been distinguished into distinct clusters this has not had a detrimental effect on the classification of the unknown CE samples as none of the unknowns are clustered in these groups.

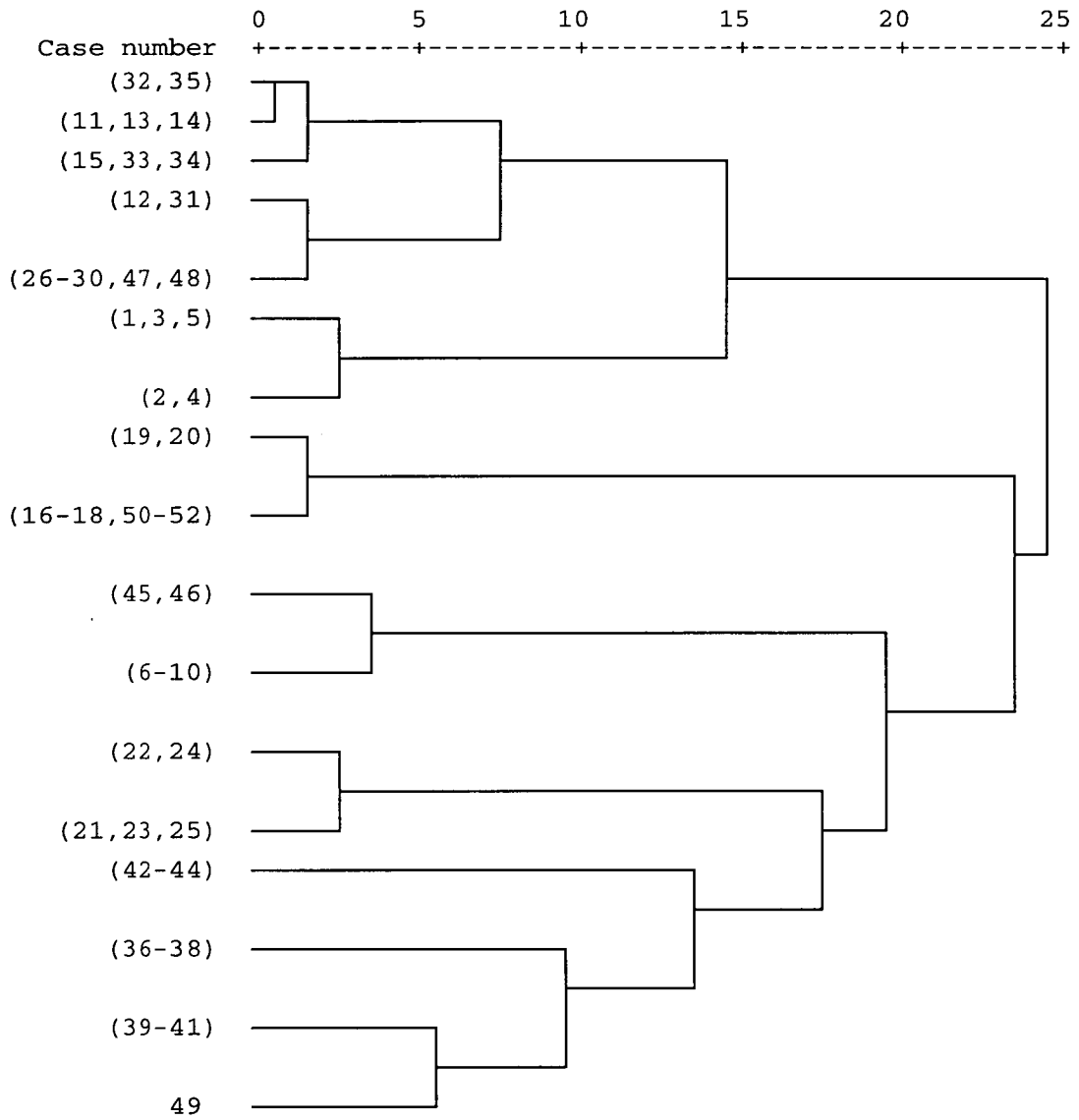


Figure 7.14 *Dendrogram of the hierarchical classification of PC scores calculated from data for known samples and unknown samples in group CE analysed by GC-FID.*

The remaining unknown samples have been separated into three groups and one single sample. The three groups representing the unknown samples (36-38, 39-41, 42-44) have not been closely associated with any of the known sample groups. Case 49 (an unknown sample) has been most closely associated with the unknown sample

cluster containing samples 39-41. As discussed in Section 6.3 the results from the interpretation of the dendrogram can be used to indicate similar samples. In the case of the three sample groups and single sample of unknown origin the distance measures show only that these samples are most similar to each other. The point at which the three samples are linked together would suggest that these samples are not from the same origin of production. Case 49 has been linked to the cluster of cases 39-41 but at a distance measure that would suggest that this is not of the same origin. The following section considers all the known and unknown samples in one large data matrix which allowed a more definite classification of the unknown cases into clusters on the basis of origin of production.

7.3.1.5 Classification of FID data of all known and unknown samples

The unknown samples were arranged into order so that the samples which were shown to be similar in the previous four sections were labelled consecutively for ease of interpretation. The samples were labelled as shown in Table 7.1. The data matrix of samples consisted of ninety one cases defined by the peak heights of thirty seven components (variables).

Table 7.1 *Key to the identity and labelling of the samples analysed by GC-FID.*

Original case numbers	Case numbers	Data label	Data type	Identity *(proposed)*
1-5	1-5	UK	Known	UK
6-10	6-10	USA	Known	USA
11-15	11-15	Japan	Known	Japan
16-20	16-20	Atochem	Known	Atochem
21-25	21-25	Hoechst	Known	Hoechst
26-30	26-30	Showa Denka	Known	Showa Denka
31-35	31-35	Ashai Glass	Known	Ashai Glass
CE 36-38	36-38	HQ	Unknown	Unknown
CE 39-41	39-41	Q	Unknown	Unknown
CE 42-44	42-44	Red	Unknown	Unknown
Q1 36-40	45-49	QF	Unknown	*USA*
Q2 36, 37	50-51			
Q3 36	52			
CE 45, 46	53-54			
Q1 41-45	55-59	QH	Unknown	*Showa Denka*
Q2 38-40	60-62			

Q3 37	63			
CE 47, 48	64			
Q1 46-49	65-68	97Q	Unknown	*UK*
Q2 41-46	69-74			
Q3 38, 39	75, 76			
CE 49	77	97Q1	Unknown	Unknown
Q2 47-51	78-82	98Q	Unknown	*Atochem*
Q3 40-45	83-88			
CE 50-52	89-91			

The data were subjected to a PCA where the eigen analysis was performed on the correlation coefficients matrix. The scree plot displaying the amount of original variation described by each component is shown in Figure 7.15. The amount of variation described by each of the components was calculated for the first twenty components and is displayed in Appendix Q. The component scores were calculated for the first ten PCs and are displayed in Appendix R.

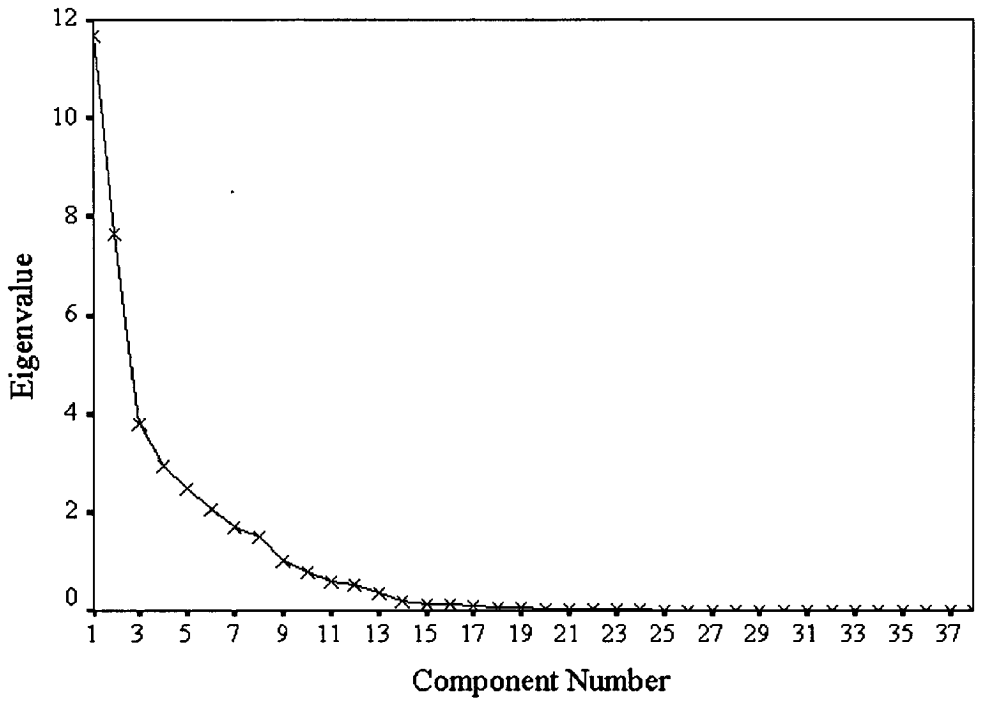


Figure 7.15 *Scree plot for PCA of all chromatographic data for samples of HFC 134a analysed by GC-FID.*

A hierarchical cluster analysis was performed using the first five PC scores (describing 75% of the original variation) to classify the samples into groups on the basis of their origins of production. The agglomeration schedule for this analysis

using Euclidean distance measures and Ward's method for cluster centre definitions is shown in Appendix S. The dendrogram drawn using these values is displayed in Figure 7.16.

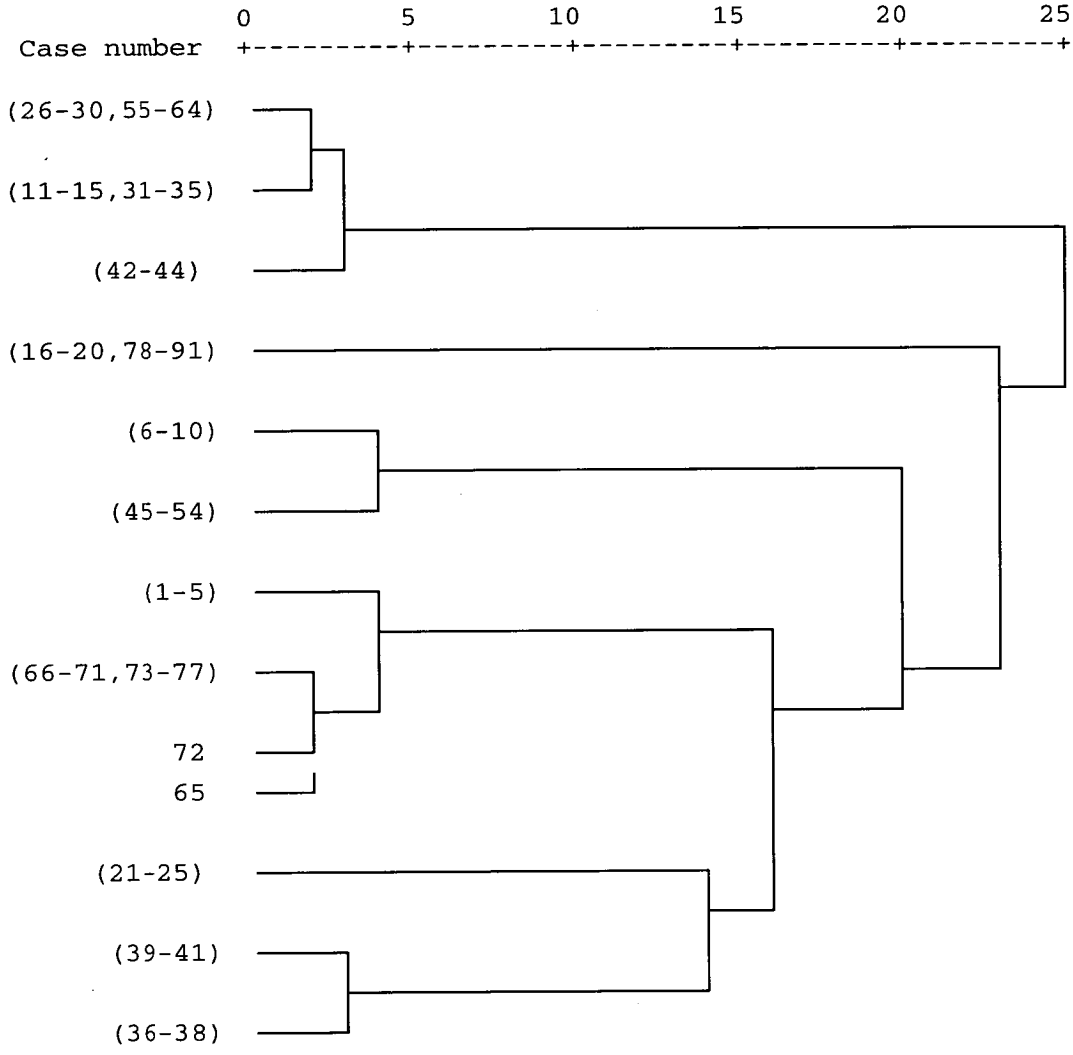


Figure 7.16 *Dendrogram of the hierarchical classification of PC scores calculated from data for all known samples and unknown samples analysed by GC-FID.*

The comparison of the classification pattern seen in Figure 7.16 and the expected classification of the samples detailed in Table 7.1 show that all the samples have been placed into the same clusters as seen when analysing the data in smaller sets.

The unknown samples 45-54 have been grouped together and closely associated with the known samples from ICI USA (6-10). The association of the samples is close enough to consider these two sample sets to originate from the same production site.

The known samples produced by Atochem (16-20) have been clustered together with unknown samples 78-91. The agglomeration schedule in Appendix S shows that the unknown samples are as closely related to the known Atochem samples as the known Atochem samples are to each other. That is, the unknown samples lie within the boundaries of variation seen within the training samples and can therefore be definitively identified as Atochem-produced samples.

The samples corresponding to Showa Denka (26-30) produced HFC 134a have been clustered together with unknown samples 55-64. The agglomeration schedule again shows that the unknown samples lie within the boundaries of intra sample variation and therefore can be considered as positively identified. The samples representing Japan and Ashai Glass have been placed into the same sample cluster and closely associated with the Showa Denka sample cluster. This follows the general trend seen in all the cluster analyses performed when using the FID data. The samples representing the unknown group 42-44 have been associated with the Showa Denka, Japan and Ashai Glass cluster. To ascertain whether these samples were closely related to this sample group a second PCA analysis was performed considering these samples in isolation.

The known samples produced by ICI Japan, Showa Denka and Ashai Glass and the unknown samples labelled RED (42-44) and QH (55-64, believed to be Showa Denka) were subjected to a PCA. The eigen analysis was performed on the correlation coefficients matrix before the PC scores were calculated for the first five components. The component scores for the first three PCs (describing 84% of the original variation) were plotted onto a three dimensional graph (Figure 7.17) to illustrate the differences between the samples. The two graphs represent the same data of PC scores viewed from two different points in space. The graph on the left hand side views the data from an angle which highlights the differences in PC2 and PC 3. The unknown samples Red (42-44) would appear in the centre of all the other samples if viewing in 2-D without considering PC1. The tilt of the axes in the PC1 dimension, however, allows the Red samples to be distinguished. The graph on the right hand side views the data from a point in space which highlights the differences

between PC1 (Red samples shown to be highly different) and PC3 (Ashai Glass samples very different).

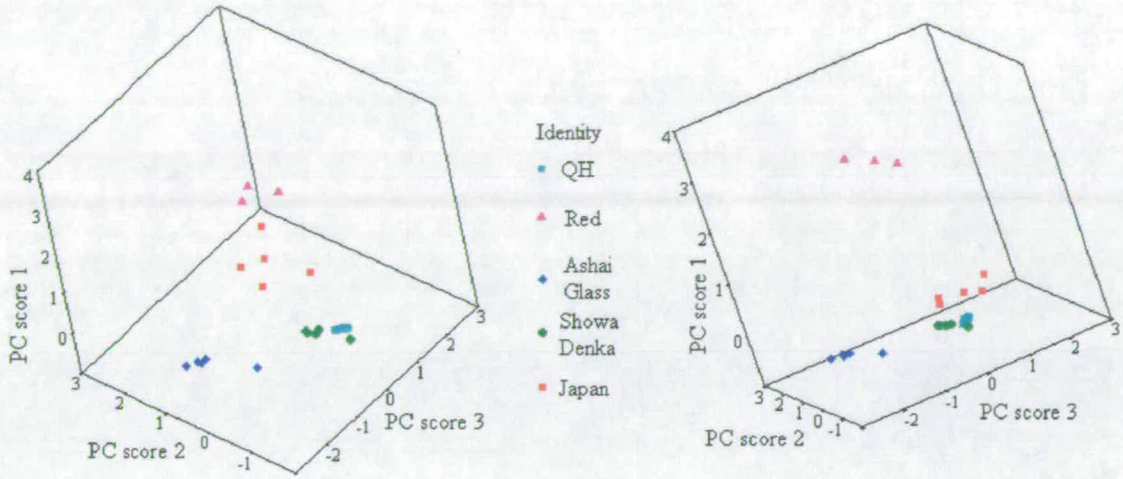


Figure 7.17 *Three dimensional plot for first three PCs for PCA of Japan, Showa Denka, Ashai Glass and a number of unknown samples.*

The graphs show that the group of unknown samples labelled QH have been placed into a cluster with the Showa Denka samples. This result matches the one seen when considering these samples in the larger data set of all samples and therefore it can be considered that these samples have been produced by Showa Denka. The samples labelled Red have been shown to be very different from the other samples considered in this group. The exact reasons for the differences can be seen by looking at the loadings applied to each PC by each variable peak. Considering the component scores singularly it is possible to see that PC1 differentiates between the unknowns Red and the rest of the samples. Considering PC1 in isolation does not allow the other samples to be correctly classified. It can therefore be concluded that by studying the loadings on PC1 (*i.e.* the contribution each peak has on the component) it will be possible to identify which peaks are causing the Red samples to be classified as different.

Figure 7.18 shows the loadings plot for PC1. The bars in Figure 7.18 show how each peak correlates to PC1 where the values range from high positive correlation (1) to high negative correlation (-1) with values approaching zero representing no correlation. The results of considering only PC1 indicates that this component can be

considered as distinguishing the Red samples from all other samples. It is therefore no surprise that the peaks showing high correlation are the ones present at high levels in only the Red sample. For example, peaks 2 and 3 are present only in the Red sample and are therefore correspondingly high. Peak 35, however, is not present in the Red samples but is seen in other samples and therefore shows a negative correlation in PC1. This method can be further applied to highlight the specific peaks and therefore components which lead to the distinguishing of the samples (see Section 8.3).

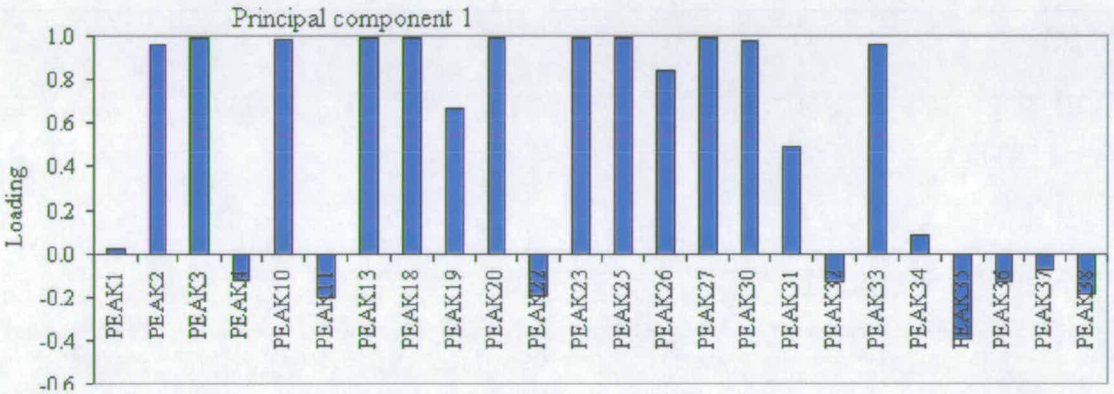


Figure 7.18 *Component loadings for PC1.*

The single unknown sample labelled 97Q1 (case 77) has been clustered with unknown samples 97Q and with the HFC 134a samples produced by ICI UK. These results show agreement with the expected results obtained from looking at the samples in smaller sets. However, the component scores for these samples are all of a similar nature, apart from PC6. This component shows a high degree of difference between the two sample groups and on the basis of this component alone the unknown samples would not be clustered together with the UK samples. The loadings for the peaks which contribute to PC6 are displayed in Figure 7.19. The peaks which are not seen in either the UK or unknown samples have not been included as these peaks will have no effect on the PC scores for the samples being considered.

The peaks which are shown to have positive correlations with PC6 are primarily those peaks (31, 33, 38) which are seen in the UK samples and not in the unknown

samples. The majority of peaks with negative correlations (2, 6, 16) are conversely seen in the unknown samples and not in the UK samples. These peak loadings are obviously affected by the other samples as well but from this inspection of the data it is possible to say that these two sets of data have generally similar chromatographic profiles but differ specifically in relation to the concentrations of component peaks 2, 6, 16, 31, 33 and 38. These peaks can then be identified and investigated in order to draw a conclusion as to the nature of the difference (Section 8.3).

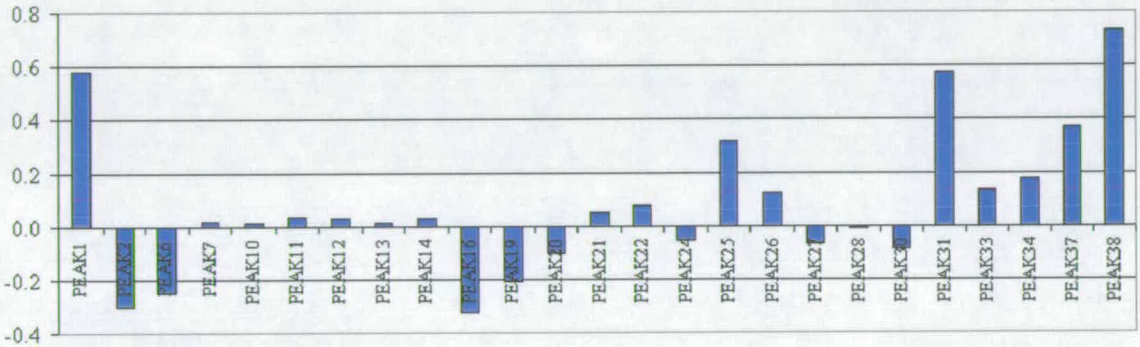


Figure 7.19 *Component loadings for peaks seen specifically in the UK and closely related Unknown samples.*

7.3.2 Classification of samples from unknown origin analysed by GC-ECD

The thirty five samples of known origin and the eight samples representing the blanks of helium and air were used to identify a series of unknown samples. The unknown samples which were previously analysed by GC-FID were similarly analysed by GC-ECD. In total there were fifty six unknown samples. As with the FID analysis, peak comparisons were made on the basis of retention time, relative retention time and mass spectral data (where available). All the peaks identified in the unknown samples had been previously seen in the known training sample set.

The unknown samples were considered in groups on the basis of where they were obtained from and their date of analysis. Each sample was profiled by GC-ECD under the standard analysis conditions to allow direct comparisons all the samples. The chromatographic data for each analysis were transferred into an SPSS spreadsheet containing all the peak height values for the training and unknown samples. The unknown data were considered in batches to simplify the initial interpretation before the data matrix as a whole was considered.

7.3.2.1 Classification of unknown sample set Q1 by GC-ECD

The Q1 sample set analysed by GC-ECD consisted of the same samples from ICI C&P as analysed by GC-FID and labelled Q1. The air and helium samples, numbered 36-39 and 40-43, respectively, were included in the GC-ECD classification data matrices. Therefore, although the unknown samples of HFC 134a appear in the same order in both the Q1 analysis by FID and ECD the absolute values are increased in the ECD analysis by eight *i.e.* FID case number 41 represents the same sample as ECD case number 49. The sample numbers were kept in the same order for both the FID and ECD analyses to allow easy comparison of results. As with the FID data the actual order of the samples in the chemometric tables does not reflect the order the samples were necessarily analysed in. The data matrix for these samples consisted of fifty seven cases defined by forty six variables.

The data matrix was investigated using a PCA to reduce the number of variables needed to describe the data. The eigen analysis was performed on the correlation coefficients matrix. The PC scores for each case were clustered using a hierarchical clustering method with squared Euclidean distance measures and Ward's method to calculate the cluster centres. The results for the cluster analysis using the first seven PCs (which describes 83% of the original variation) are displayed in the dendrogram in Figure 7.20.

The classification of the samples has placed a Japan sample (case 11) into the cluster containing the helium samples which is the same misclassification as seen when considering the training samples in isolation. Similarly the remaining Japan samples have been classified into a group containing an Ashai Glass sample. The misclassifications of these samples was common when considering the data previously as seen in Section 6.1.2.4-6 and are not considered in this section as they have no effect on the classification of the unknown samples.

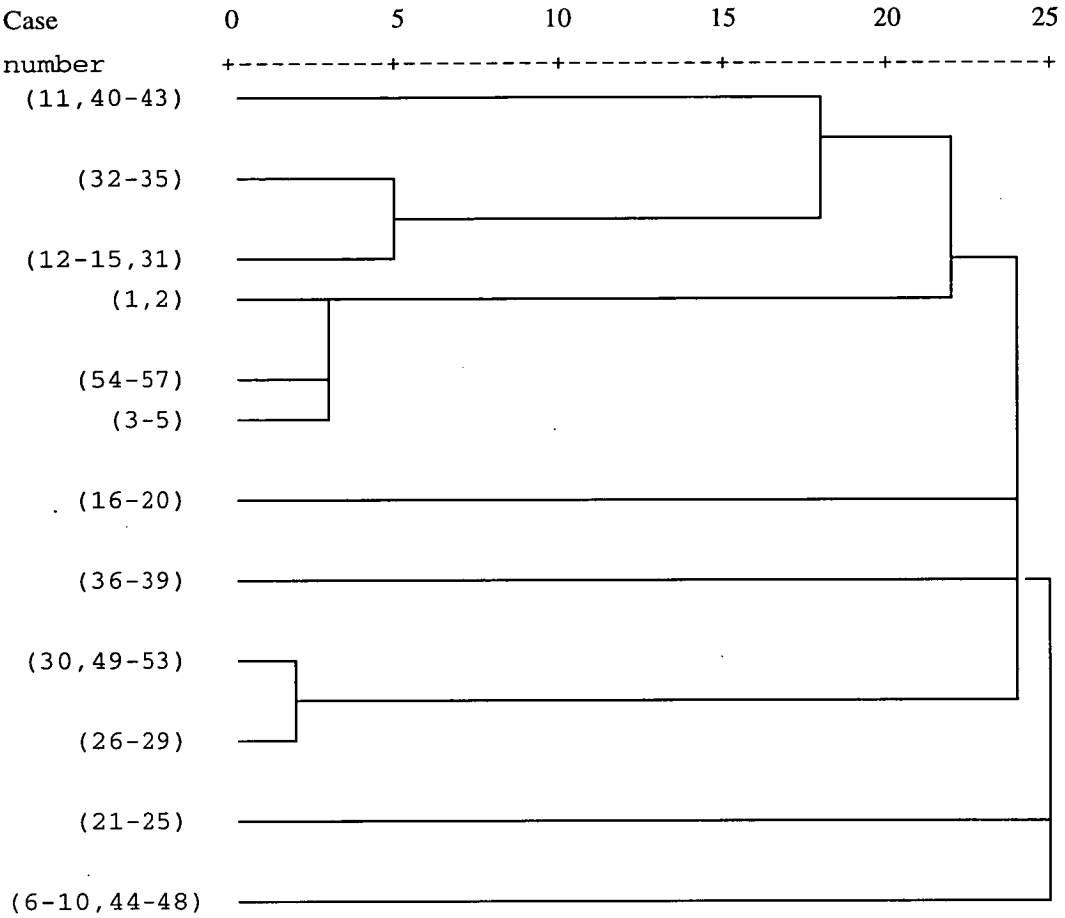


Figure 7.20 *Dendrogram of the hierarchical classification of PC scores calculated from data for known samples and unknown samples in group Q1 analysed by GC-ECD.*

The Q1 unknown samples 54-57 (46-49 in the FID analysis) have been classified into a sample cluster containing the UK known samples. The variation of the unknown samples from the known samples lies within the intra sample variation for the known samples. Therefore, the unknown samples can be considered as belonging to the same origin of production as the UK samples.

The samples corresponding to Atochem (16-20), Hoechst (21-25) and air (36-39) have all been clustered correctly into separate clusters which display a high level of dissimilarity from all other samples as they are only clustered in the final stages of the agglomeration schedule.

The unknown samples 49-53 have been clustered into the same group as the known samples from Showa Denka (26-30). The level of similarity is highlighted by case 30 (Showa Denka) being clustered with the unknowns before the other Showa Denka samples. Finally the cases representing USA have been closely clustered into a single group with the unknown samples 44-48.

The classification of this first set of samples on the basis of the ECD data has clustered the unknown samples into exactly the same clusters as seen with FID data set. That is, although the actual data sets are defined by very different peak heights they have produced the same results using the two detection methods. This adds an extra level of confidence to the assignment of sample identities as both methods show that the measurement of the by-products of production allows the identification of samples into the same groups.

7.3.2.2 Classification of unknown sample set Q2 by GC-ECD

The sample set Q2 analysed by GC-ECD contained the same samples as those analysed by GC-FID in section 7.3.1.2. The samples were all obtained from ICI C&P and analysed during the second year of this work. This sample set of unknowns consisted of sixteen samples described by forty six variables. The chromatographic profiles were aligned with the training samples of known origin on the basis of retention time, RRT and mass spectral identification (where available).

The data for the unknown samples set Q2 were entered into the data matrix containing the results for the GC-ECD analysis of the known samples. The data matrix was subjected to a PCA where the eigen values were extracted from the correlation coefficients matrix. The first ten PC scores were calculated for each case and used to describe the data. A hierarchical cluster analysis was performed on the first seven PC scores for the samples using Euclidean distance measures with Ward's method to describe cluster centres. The classification results for this analysis are shown in Figure 7.21 in the form of a dendrogram.

The samples representing Ashai Glass (31-35), Japan (11-15) and Helium (40-43) have not been fully resolved into distinct samples groups on the basis of their origin of production. However, as none of the unknown samples appear in these groups full resolution is not required at this stage of the analysis. The samples representing HFC 134a from Hoechst (21-25) and the blanks of Air (36-39) have been placed into distinct groups containing none of the unknown samples.

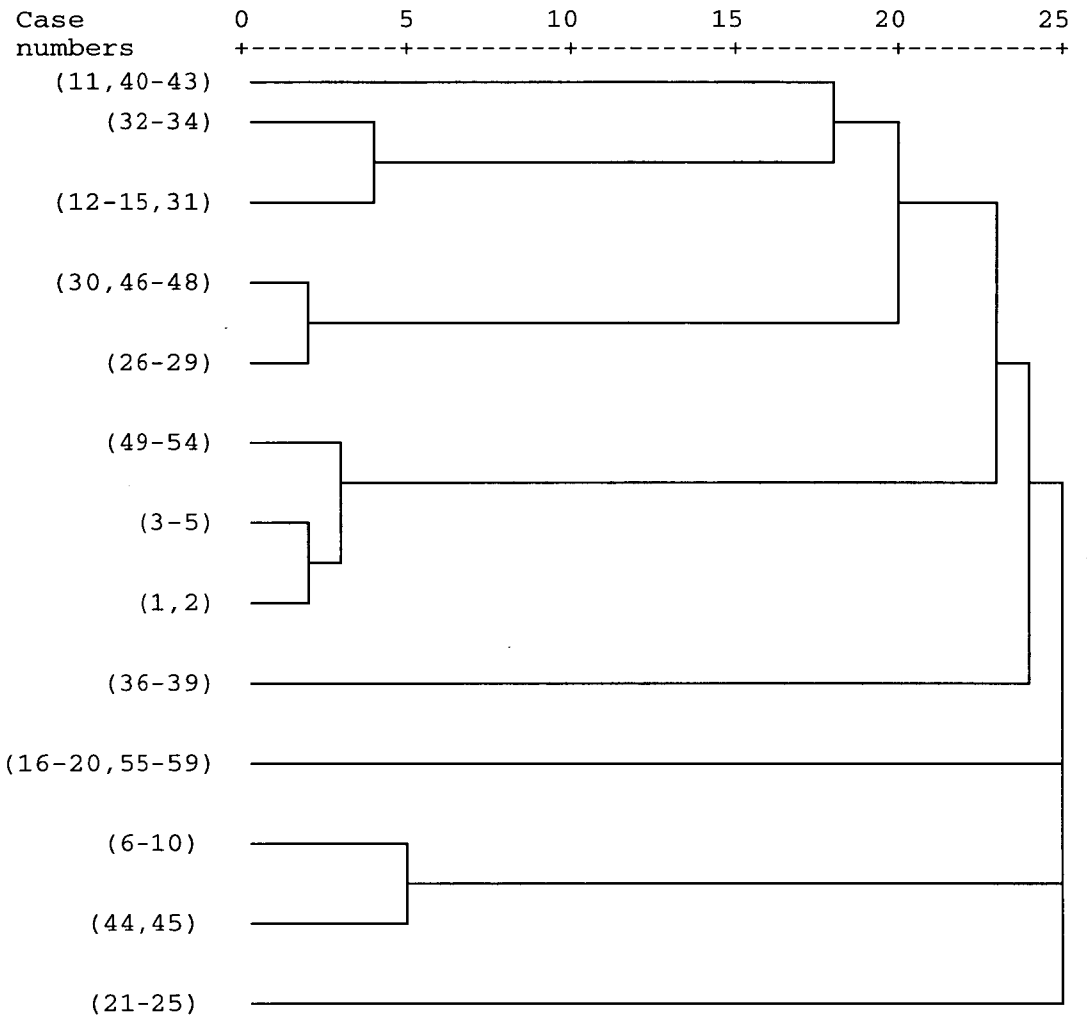


Figure 7.21 *Dendrogram of the hierarchical classification of PC scores calculated from data for known samples and unknown samples in group Q2 analysed by GC-ECD.*

The unknown Q2 samples 46-48 have been clustered into a group containing samples produced by Showa Denka. The re-scaled distance at which these samples have been associated shows that these samples fall within the area of intra sample variability and can therefore be considered as being from the same origin of production.

Similarly, Q2 unknowns 55-59 have been clustered with the samples from Atochem and can therefore be considered as being from the same origin of production.

The unknown Q2 samples 49-54 have been closely associated with the UK samples (1-5) and the unknown samples 44 and 45 have been closely associated with the USA samples (6-10). In both instances the unknowns lie outside the intra sample variation of the known origin samples but are still closely associated with the known samples. This level of similarity implies that the samples are from the same origin of production although the fact that they have been placed into a distinct cluster implies that there are real differences between the samples; these differences are investigated further in section 8.3.3.

The classification of the unknown sample group Q2 by ECD has produced the same identities for the samples as produced by the FID analysis. This shows again, as with sample group Q1, that the proposed sample identities are correct, since the two distinct methods result in the same conclusions as to the origins of production of the unknown sample set. It should also be noted that chemometric analysis of the ECD data for the unknown samples did not require a second PCA to clearly discriminate the Showa Denka samples (both known and unknown) from the Japan and Ashai Glass samples.

7.3.2.3 Classification of unknown sample set Q3 by GC-ECD

The sample set Q3 contained ten samples of unknown origin of production obtained from ICI C&P. The samples were analysed by GC-ECD and the chromatographic profiles compared with the samples of known origin. Peak identification was made on the basis of RT, RRT and mass spectral information (where available).

The data matrix of known and Q3 samples consisted of fifty three samples defined by forty six variables. The PC scores were calculated from the PCA based on the correlation coefficients matrix and used to classify the samples by performing a hierarchical cluster analysis using Euclidean distance measures and Ward's method

for cluster centre definitions. The dendrogram drawn from the re-scaled cluster distances is shown in Figure 7.22.

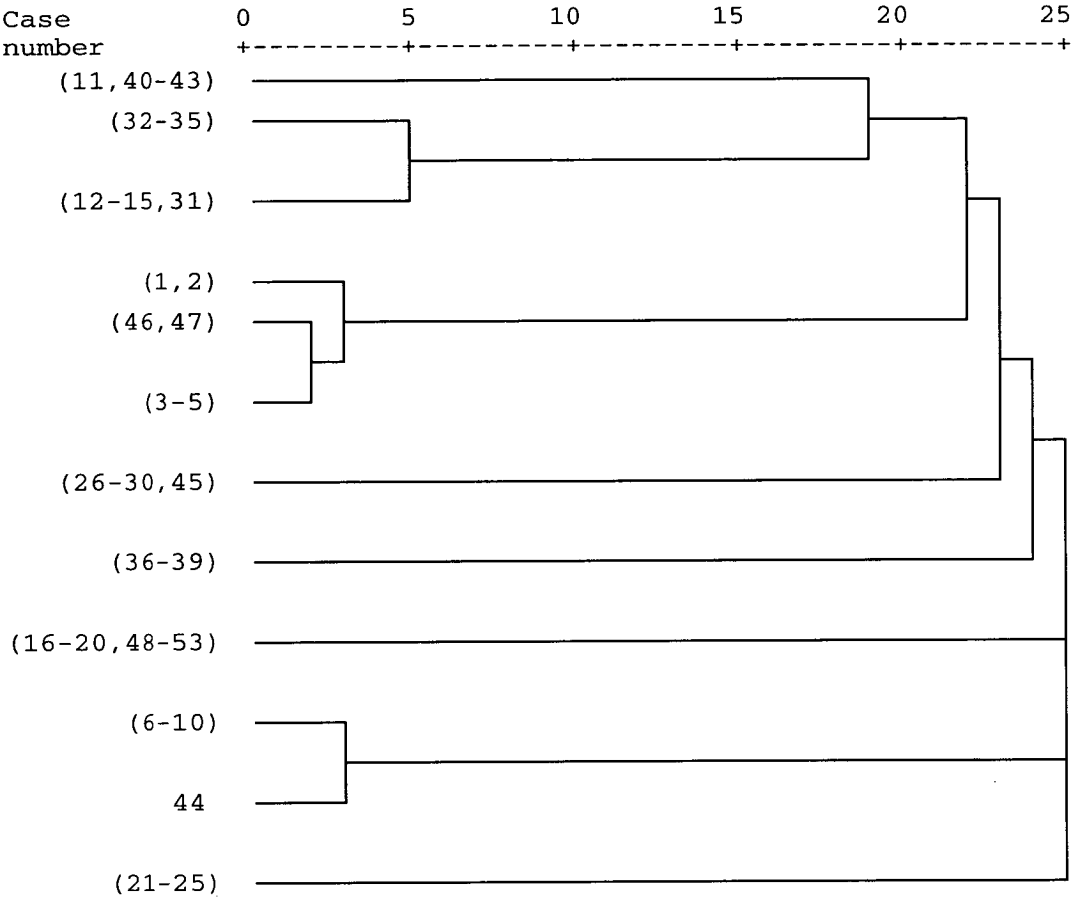


Figure 7.22 *Dendrogram of the hierarchical classification of PC scores calculated from data for known samples and unknown samples in group Q3 analysed by GC-ECD.*

The classification pattern for the Japan (11-15), Ashai Glass (31-35) and Helium blank (40-43) samples matches that seen in the previous two analyses of the ECD data looking at the Q1 and Q2 unknown samples (sections 7.3.2.1 and 7.3.2.2). Again, no samples of unknown origin were associated with these misclassified groups and therefore no further chemometric analyses were required to correctly classify these samples.

Nine of the unknown samples were placed into clusters within the limits of intra cluster variation. These were cases 46 and 47 clustered with UK samples (1-5), case 45 clustered with Showa Denka samples (26-30) and cases 48-53 clustered with Atochem samples (16-20). These unknown samples can be considered as being from

the same origin of production as the known samples with which they were clustered. Case 44 was closely associated with samples from the USA (6-10) at a distance level which shows this sample to be only slightly different from the known USA chromatographic profile as determined by ECD.

The classification of all the unknown samples from Q3 showed close agreement with the identifications stated in Section 7.3.1.3, again showing that profiling using either detection method produces the same and hence corroborated these answers as to the origins of the unknown samples.

7.3.2.4 Classification of unknown sample set CE by GC-ECD

The unknown sample set CE contained the samples obtained from the Chemical Engineering Department of Edinburgh University. This sample set consisted of seventeen samples described by forty six variables. The chromatographic data for these samples were entered into a data matrix containing the results for the ECD analysis of the known samples. This data matrix was subjected to a PCA using the correlation coefficients matrix followed by a hierarchical cluster analysis of the PC scores for the first eight PC. This increase in the number of PCs used to define the data reflects the increase in the variation of the samples when looking at the CE sample set. That is, the results of the analysis using seven PC scores to describe the data did not produce the optimum separation of the samples. The results of the chemometric analysis of the known and unknown samples CE are displayed as a dendrogram in Figure 7.23.

The overall pattern of the dendrogram in Figure 7.23 has changed from that previously seen when looking at the ECD data (Figures 7.20-22). The classification of the sample groups is still the same but the linking of clusters representing different origins of production has changed.

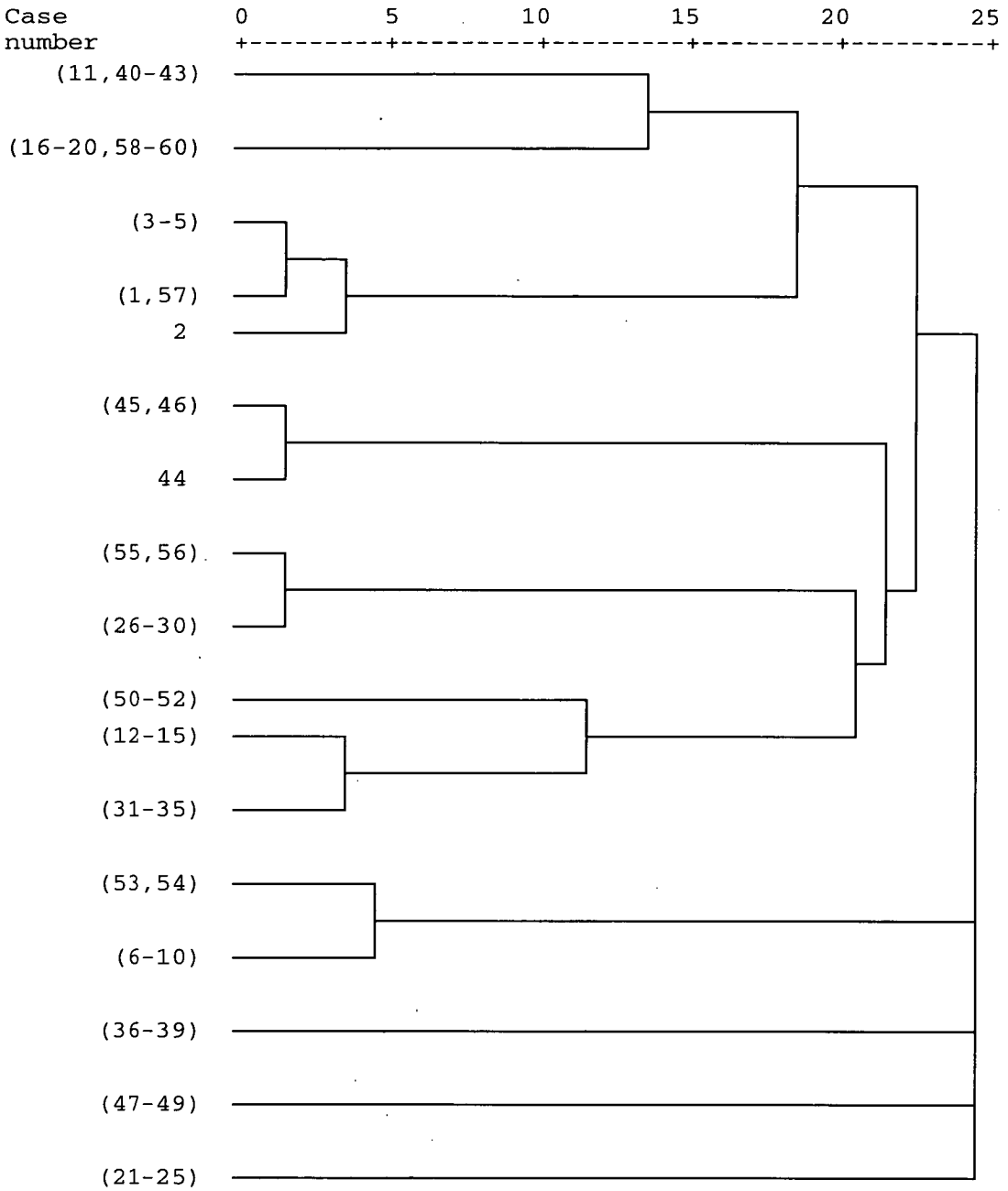


Figure 7.23 *Dendrogram of the hierarchical classification of PC scores calculated from data for known samples and unknown samples in group CE analysed by GC-ECD.*

The unknown samples 58-60 have been classified into the same group as the known samples from Atochem (16-20). The unknown sample 57 has been classified in the group of origin UK (1-5). These unknowns have been classified within the intra variation of the known samples and can therefore be considered as definitively identified.

Four of the unknown samples have been closely associated with clusters representing samples of known origin. These were cases 55 and 56 linked with Showa Denka (26-30) and cases 53 and 54 which were closely linked to USA (6-10). These unknown samples can be considered as being highly similar to the groups they have been associated with but not positively identified.

The remaining unknown samples have been clustered into three distinct groups, cases 44-46, cases 47-49 and cases 50-52. The cluster containing cases 50-52 has been associated with the cluster containing the Japan and Ashai Glass samples. However, the distance at which these samples have been clustered together suggests that this reflects only a loose similarity and not a real connection between the samples.

7.3.2.5 Classification of ECD data of all known and unknown samples

The data representing all the known samples and the four groups of unknown samples were considered as one data matrix. The unknown samples were arranged into order so that the samples which were considered to be similar, as shown in the previous four sections, were labelled consecutively for ease of interpretation. The samples were labelled as shown in Table 7.2. The data matrix of samples consisted of ninety nine cases defined by the peak heights of forty six components (variables).

Table 7.2 *Key to the identity and labelling of the samples analysed by GC-ECD.*

Original case numbers	Case numbers	Data label	Data type	Identity *(proposed)*
1-5	1-5	UK	Known	UK
6-10	6-10	USA	Known	USA
11-15	11-15	Japan	Known	Japan
16-20	16-20	Atochem	Known	Atochem
21-25	21-25	Hoechst	Known	Hoechst
26-30	26-30	Showa Denka	Known	Showa Denka
31-35	31-35	Ashai Glass	Known	Ashai Glass
36-39	36-39	Air	Known	Air
40-43	40-43	Helium	Known	Helium
CE 44-46	44-46	HQ	Unknown	Unknown
CE 47-49	47-49	Q	Unknown	Unknown
CE 50-52	50-52	Red	Unknown	Unknown
Q1 44-48	53-57			

Q2 44, 45	58, 59	QF	Unknown	*USA*
Q3 44	60			
CE 53, 54	61, 62			
Q1 49-53	63-67	QH	Unknown	*Showa Denka*
Q2 46-48	68-70			
Q3 45	71			
CE 55, 56	72, 73			
Q1 54-57	73-76	97Q	Unknown	*UK*
Q2 49-54	77-82			
Q3 46, 47	83, 84			
CE 57	85			
Q2 55-59	86-90	98Q	Unknown	*Atochem*
Q3 48-53	91-96			
CE 58-60	97-99			

The data were subjected to a PCA where the eigen analysis was performed on the correlation coefficients. Figure 7.24 shows the scree plot displaying the amount of original variation described by each component. The points on the scree plot show the way in which the variation in the data is spread over more of the components; therefore more components need to be considered in order to describe the majority of original variation within the system. The component scores were calculated for the first twenty PC and are displayed in Appendix T.

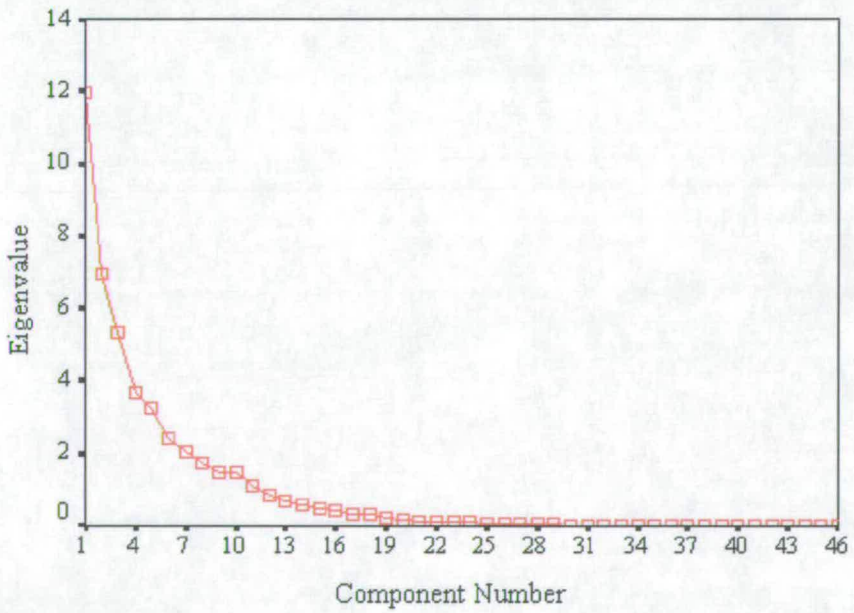


Figure 7.24 *Scree plot for PCA of all chromatographic data for samples of HFC 134a analysed by GC-ECD.*

A hierarchical cluster analysis was performed on the first ten component scores (Appendix U) using squared Euclidean distance measure with Ward's method to describe the cluster centres. The agglomeration schedule for this analysis is displayed in Appendix V with these results shown as a dendrogram of linkages in Figure 7.25.

The samples representing the UK (1-5) have been closely linked with the unknown samples 73-85. The unknown samples have been linked within the intra sample variation shown between the samples known to have been produced by ICI UK and can therefore be considered as being of the same origin. This conclusion matches that reached in the FID analysis of the same samples (Section 7.3.1.5) and can therefore be considered as a definite identification.

The cases representing the unknown samples 86-99 have all been closely clustered with the samples produced by Atochem (16-20). The level of linkage shown in the agglomeration schedule shows that the some of the unknown samples are as similar to the known samples as the known samples are to each other. These results match those seen in the FID data analysis and can therefore be considered as showing that these samples originated from Atochem.

The unknown samples (63-72) which were provisionally identified as being of the same origin as the Showa Denka samples (26-30) are shown to fall into the same cluster as the samples from Showa Denka when considering the data as a whole. The level of similarity, as defined by the cluster distance measures, and the agreement with the results in the FID analysis show these unknown samples to be of this origin of production.

The samples corresponding to production by ICI USA (6-10) have been clustered together and then linked with a group of unknown samples (53-62) which were shown in the previous four sections to be of a higher similarity. Closer inspection of the data shows that these two sets of samples show similarities in certain areas and

differences in others. That is, by considering only selected PCs the two sample sets do appear similar. However, by considering more PCs, and therefore more of the variation in the samples, these groups are not being classified as closely associated This implies that they are either not from the same source or that there are inherent variations seen in samples from the same source.

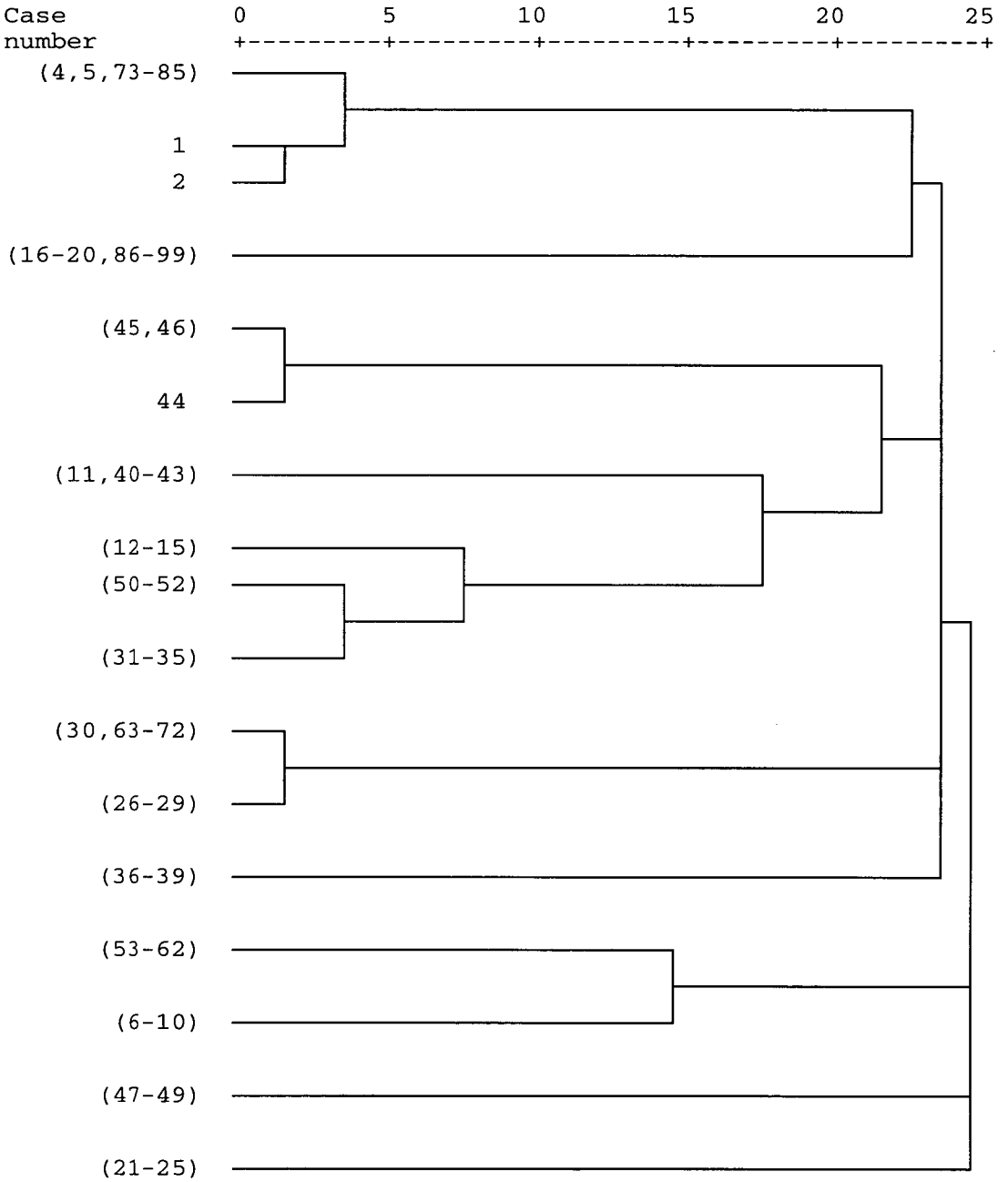


Figure 7.25 *Dendrogram of the hierarchical classification of PC scores calculated from data for all known and unknown samples analysed by GC-ECD.*

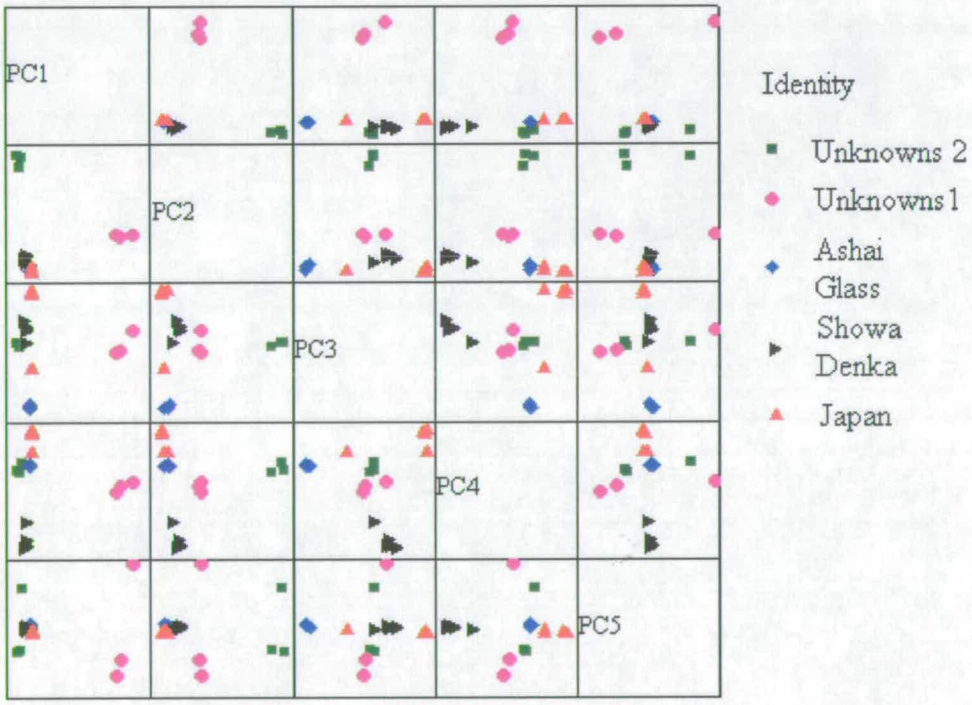
The cases representing the blank analysis of air have been clustered into a distinct group and positioned separately from all other samples except at the final stages of the agglomeration schedule, showing these samples to be very different from all the other samples analysed.

The unknown samples 50-52 have been placed into a single cluster which has been closely associated with the Ashai Glass samples (31-35). These samples have then all been linked to the Japan cluster (12-15). This cluster was then linked to the Helium samples (containing a misclassified Japan (11) sample). This cluster was then finally linked with the unknown samples of 44-46 before all being linked to the other samples in the agglomeration schedule. These samples all have the general trend of containing fewer peaks of notable size when compared with the other samples. The classification pattern seen for these samples is highly reflective of this lack of peaks. Therefore these samples have been classified as being similar when considering all the data, primarily because of the fact that they do not contain peaks, rather than on the basis of the peaks they do contain. It is acceptable to consider these samples as being dissimilar to the other samples *e.g.* Hoechst samples, as they contain few similar peaks at the same levels but it is not as acceptable to consider these samples as the same due to a lack of information to discriminate them. It was therefore necessary to further investigate these samples in isolation to see if the clustering pattern remained when only considering the peaks present in this smaller sub-set of the data.

This reduced data set consisted of twenty cases defined by forty variables. The data were subjected to a PCA and the component scores for the first five components were calculated and used to define the samples. These component scores were plotted onto a Draftsman's plot in order to illustrate the differences between the samples (Figure 7.26).

It is apparent from the figure that the samples have been shown to be different. Considering PC1 in isolation it is clear that the group of samples labelled Unknowns

1 is highly different from all the other samples. As with the whole data set the PCs each appear, in general, to highlight one set of samples as being different from the others. PC1 differentiates Unknowns 1 from the rest, PC 2 differentiates Unknowns 2 from the other samples, PC3 differentiates Japan and Ashai Glass from the others and PC4 highlights Showa Denka as being specifically different from the others. However, although on the large scale it appears that PC1 only separates Unknowns 1 from the rest, by enlarging this scale it is possible to see that all the other samples are distinguished from each other as well (Figure 7.27). Extending this effect when looking at all the data would lead to the conclusion that for the majority of the PCs which are used to describe the data, the samples with the lower levels of components will often be described by values which although distinct from one another appear close together when considered with the other samples. This means that it is important not to simply consider the dendrogram distance measures without



consulting the raw data or performing further PCA on those samples which appear similar.

Figure 7.26 *Draftsman's plot of first five PC for the PCA of selected sub-set of ECD data for known and unknown samples.*

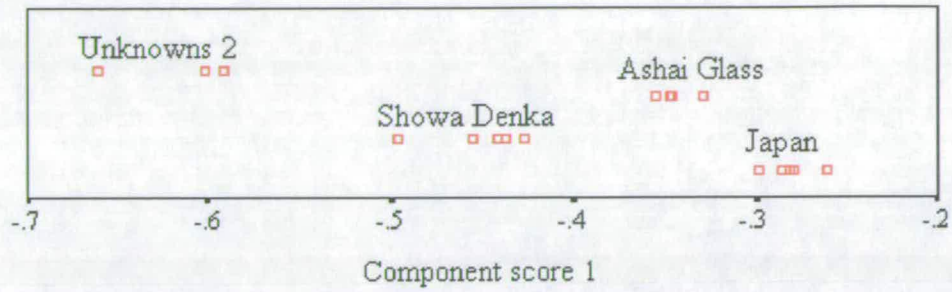


Figure 7.27 *Enlarged scale plot of first PC for the PCA of selected sub-set of ECD data for known and unknown samples.*

The Hoechst known samples (21-25) and the unknown samples labelled cases 47-49 have been placed into distinct clusters and are only clustered with other samples in the final stages of the agglomeration schedule. However, it should be noted that when considering only PC1 and PC2 the Hoechst and unknown samples 47-49 do show a high level of similarity. The consideration of further components leads to the two samples sets being classified as very different from each other, as seen with the USA (6-10) samples and unknown cases 53-62.

7.4 Chemical implications of the classification patterns

The previous sections have shown that samples of HFC 134a from different origins of production can be separated and classified into groups on the basis of similarities and differences between their chromatographic profiles[5,6]. These chromatographic profiles differ on the basis of the different samples relative contents of different by products of synthesis. These by-products of synthesis can include compounds arising from impurities in the reaction materials, non specific reactions within the synthesis process or compounds contaminating the sample post synthesis. The levels of these components are also highly reliant on the amount of post synthesis purification and the subsequent specificity for each component these purification steps have. The following section discusses the characterisation of the samples on a general basis highlighting the trends in the sample groups.

7.4.1 Gross trends within the sample classification

The classification patterns produced by the FID and the ECD chromatographic analyses both displayed the same general trends, even though the actual

chromatographic traces were markedly different. The major classification trend can be directly linked to the amount of information contained within a set of sample profiles. That is, the samples which contained the most peaks at the highest levels were shown to be different from all the other samples.

A number of the samples had low levels of by products present and therefore peaks were often below the limits of detection. The samples representing Japan, Ashai Glass and Showa Denka were consistently the most difficult to classify into distinct groups. This was because of the level of similarity these samples displayed when considering the full range of potential by-products. As the majority of the peaks for these compounds were below the limit of detection these samples were classified similar on not on the basis of the peaks they contained but on the peaks they did not contain. This is apparent when performing the hierarchical cluster analysis of the samples on the basis of the PC scores obtained from looking at all the data. The samples representing UK, USA, Atochem and Hoechst all have larger numbers of peaks at higher levels than the Japan, Ashai Glass and Showa Denka samples. Therefore, when considering the whole data set the samples containing fewer peaks are classified closer together which leads to misclassification when using the clustering techniques employed. However, if these samples are considered in isolation by performing a PCA and clustering only on the Japan, Ashai Glass and Showa Denka samples they are classified into discrete sample groups. Similarly, as the ECD data provides more peak information because of its lower limits of detection for a number of different components the trend is to obtain better classifications using the ECD data rather than the FID data.

The number of peaks present in each sample by which they were classified had a major effect on the ability of the technique to correctly classify samples[7]. Therefore the amount of post production purification will dramatically alter the classification of samples. The samples analysed in this work were all of an industrial grade and therefore contained what could be considered a reasonable number of peaks to be classified by. If the grade of the sample was improved and hence the level of by

products reduced the number of peaks on which the classification was based be would lowered. It would therefore be apparent from the work discussed here that to classify different grades of HFCs would lead to the classification of the highest purity samples into similar groups. This would then require a further analysis of these pure samples in isolation to allow their separation into clusters. Obviously taking this argument through to its conclusion at a certain point the levels of by products would be reduced to such a level that only a few peaks would be detectable rendering a multi variate analysis pointless.

References

[1] Brown, J.S., Crop Sci., **31**, V4, p915-922, (1991)

[2] Bicchi, C.P, Panero, O.M., Pellegrino, G.M. and Vanni, A.C., J. Agric. Food Chem., **45**, p4680-4686, (1997)

[3] Malmquist, G. and Danielson, R., *J. Chrom. A*, **687**, p71-88, (1994)

[4] Wold, S., Esbensen, K. and Geladi, P., *Chemom. Intell. Lab Syst.*, **2**, p37-52, (1987)

[5] Lamparczyk, H., Ochocka, R.J., Grzybowski, J., Halkiewicz, J. and Radecki, A., *Oil and Chem. Pollution*, **6** p177-193, (1990)

[6] Myors, R., Wells, R.J., Skopec, S.V., Crisp, P., Iavetz, R., Skopec, Z., Ekangaki, A. and Robertson, J., *Anal. Comms*, **35**, N12, p403-410, (1998)

[7] Stout, S.A., *J. Anal. Appl. Pyrol.*, **18**, p277,292, (1991)

Chapter 8

Summary

8.0 Summary and discussion of work performed

The work performed in this project had three main goals. The first was to develop and test an analytical method which would provide reliable descriptive data that was representative of the different samples of HFC 134a studied. The second was to take this descriptive data and apply chemometric analysis techniques in order to classify samples of HFC 134a into groups on the basis of similarities between these samples. The final part of the work was to investigate the data and try to ascertain the reasons for the separation of the samples into different clusters especially in relation to the route and method of synthesis of each different sample. The previous chapters have outlined the steps and taken to achieve these goals and the results of the analyses performed. This chapter summarises, draws conclusions and details what further studies could be performed to progress the work detailed in this thesis.

8.1 Importance and strategy for the classification of HFC 134a

The introduction to this thesis detailed the background to the implementation of the use of HFC 134a as a major industrially synthesised chemical[1]. The environmental problems associated with the use of CFCs were discussed which highlighted the need for replacement compounds, a major one being HFC 134a. The reduced environmental impact of HFC 134a with respect to ozone depletion and global warming and its thermodynamic properties have led to its increase in production as one of the major replacements for CFCs[2]. The increase in use to date and that forecasted shows this compound to be of major industrial importance.

The need to monitor the production of HFC 134a and then to track the final product have been highlighted. The application of a fully automated method for this analysis and subsequent interpretation of these analysis results have been shown to be required. Therefore the implementation of the ideas and methods discussed and developed in this work have been shown to have good standing.

The routes of synthesis of the different samples of HFC 134a were presented and discussed[3]. Unfortunately, because of the industrial importance of this compound

and the secrecy in which HFC 134a is produced the fine details of the synthesis routes were not known. In particular, details on the specific catalysts, temperatures and clean-up steps employed in the different synthetic routes were not known. This leads to a reduction in the amount of direct information that can be gleaned from the separation patterns observed. However general conclusions on the reasons for the different compositions of by-products have been presented.

Classification of the different samples of HFC 134a on the basis of their content of contaminating by-products was discussed. GC was chosen as the method for measuring these by-products as it allowed the separation and quantification of each of the different by-products present in the samples in one analytical process[4]. The general theory of GC analysis was discussed with particular reference to how to obtain the most accurate and precise results. The text detailed the equipment parameters which needed to be considered, and in many cases monitored, in order to obtain the best quality data.

The choice of detection methods for the by-products of synthesis of HFC 134a were flame ionisation, electron capture and mass spectrometry. The method of action of each detection system was discussed paying particular attention to the mass spectrometer as this provided the most important information for the identification of by-products. The true potential of the mass spectrometer as an analytical tool was not fully realised in this work because of equipment limitations and restraints on access time. The relative pros and cons of each detection technique were discussed in reference to their ability to analyse and provide information on the levels of by-products present. Of particular interest were the extremely high levels of sensitivity achievable using the ECD detector when looking at by-products containing chlorine and bromine[5].

The final sections in the introduction dealt with the processing of the signals obtained from the different detectors. The concepts of background noise, which is present in all the chromatographic traces, baseline drift, which appears under certain conditions,

and the impact these two parameters have upon the identification and definition of what is and is not a chromatographic peak were detailed. The concepts of limit of detection and limit of quantification were discussed and applied to the data obtained in the later analysis of the samples of HFC 134a studied.

Chapter 2 introduced some of the data analysis and classification methods which fall under the description of chemometrics. The term chemometrics was shown to include a multitude of different mathematical processes[6] which can be applied to chemical data of which only a small number could be considered in this work. The methods that were to be applied to the classification of the chromatographic data were introduced and discussed paying particular attention to PCA which was to be used as the major exploratory tool in the work performed.

The principles behind the classification of data were also discussed, in particular the use of clustering techniques. These were presented as the methods that would allow the classification of samples into distinct groups on the basis of their chromatographic profiles. The unsupervised pattern recognition method of hierarchical clustering was discussed and showed how this method of analysis can be applied to solve classification problems. A second classification method, K-means clustering, which relied on a more subjective analysis of the data was detailed. The K-means clustering method required a certain amount of prior knowledge of the samples to be applied as the number of cluster centres needed to be stated before the analysis was performed.

The use of a training set of data samples to form the basis of a classification method was introduced. This showed that by analysing a series of known samples and using this data as a training set, classification methods could be developed. These were then tested by classifying new samples of known origin in order to demonstrate that the methods of classification worked. These tested methods were than applied to identify samples blind to the operator and attempt to classify samples of unknown origins.

8.2 Experimental design

The discussions in Chapter 2 dealing with the analysis of the data, by whichever method, all have a common theme, which refers to the quality of chemical data. Many publications compare and contrast different methods of sample characterisation but all these methods rely heavily on the quality of the data and the experimental design therein. It is pointless trying to analyse the data using convoluted and often complex methods if the underlying quality of the data is poor. That is, it is impossible to tell the difference between two groups of samples if the levels of intra sample variation exceeds the levels of inter sample variation. It is therefore vitally important to optimise all conditions of analysis to attempt to minimise any variation within the system. Under ideal circumstances, the majority of the variation between samples will be because of true differences and not because of a systematic error. It should be noted that random error will always be present.

8.2.1. Optimisation of experimental conditions

The methods of sample handling using gas tight systems and the frequent 'cleaning' of vessels using either vacuuming and/or washes with helium were employed to minimise the levels of potential contamination from external sources. A glass vacuum chamber was designed and built in-house in order to dilute and mix gases at specific levels. This allowed response curves and mixed sample analyses to be performed with a high level of accuracy and purity.

The building and testing of a gas injection system provided the means to reproducibly analyse fixed quantities of gaseous samples. Extensive testing of the gas injection system determined the optimum volumes required for operation and demonstrated that the correct operation of the injector would allow accurate and precise loading of set volumes onto the GC column[7].

The choice of GC column to perform the separation of by-products required to facilitate the classification of the samples was based on literature searches and personal recommendations from colleagues at ICI. Literature searches showed that

no single GC column provided the separation required without exhibiting some drawbacks[4]. The time and financial constraints applied to this work prevented a full evaluation of this wide range of GC columns discussed in the literature. It was therefore decided to use a PLOT alumina column deactivated with $NaSO_4$ as this had been shown to provide separation of the low molecular weight halocarbons expected to be encountered in this work[8]. Changes in the chromatographic profile because of the potential dehydrochlorination of a number of the lower molecular weight halocarbons[9] were monitored. Although no direct instances of this were observed, it is impossible to ascertain, from the work performed here if this effect did lead to the reduction or elimination of certain peaks. However, it is possible to conclude that if dehydrochlorination was occurring, the effect was consistent and therefore would not impact on the overall classification of the samples of HFC 134a.

The equipment used for the detection of the by-products was tested and optimised in order to give the most accurate and precise results. The relatively simple operation and mechanism of action of the FID for the detection of by-products meant that only the operating temperature was investigated. This investigation showed that provided a minimum operating temperature was exceeded (200°C) the detector would operate to its optimum. The ECD operating parameters of both nitrogen make up gas flow and operating temperature had a major effect on the sensitivity of the detector to the by products analysed. However, in both cases varying from the recommended manufacturers operating conditions would lead to long term deterioration of detector performance. Therefore the operating conditions were optimised within the recommended limits (30mls.min^{-1} nitrogen flow, temperature of 350°C). The mass spectrometers used in this work were regularly calibrated and the operating conditions tuned in order to produce a maximal response for HFC 134a. Optimising the conditions using other standard halocarbons produced similar operating conditions showing the optimised conditions for HFC 134a would be applicable for all the halocarbons encountered in this work.

The quality of the signal obtained from the different detectors was investigated by measuring levels of noise and baseline drift obtained from the analysis of HFC 134a by both FID and ECD. In each case the levels of noise and drift remained constant and below the levels acceptable to the analysis criteria outlined in Section 3.6 of this work. The performance of each detector was monitored by the analysis of standards at regular intervals, the Shewhart charts showed consistent performance of each detector. The MS was not so thoroughly investigated as the data from this method of analysis would be used only in a qualitative sense.

The measured noise and drift levels were used in the formation of a definition for what was and was not classified as a peak[10]. Limitations in the software used to collect the data from the GC meant peak heights were used to characterise the peak responses seen in the chromatographic traces. This led to tighter constraints being applied to the quality of the chromatographic columns used as changes in the column performance would have direct effects on the measured peak response. The temperature separation profile was also kept constant throughout the course of this work to avoid changes in peak heights because of chromatographic effects.

8.2.2 Identification of by-product peaks

The identification of the by-products present in each chromatogram was necessary to allow the application of the multi variate analysis and subsequent classification[11]. The actual identity of the by-product was not required for classification purposes but was obviously required when trying to assign some chemical meaning to the differences seen between the samples. Correct classification of the samples required the assignment of peak identities between the chromatographic traces. That is, provided that Peak 12 was the same in each chromatographic trace then to actually identify Peak 12 was not required. The initial assignment of peak identity was performed on relative retention times[12] to the leading edge of the peak representing HFC 134a. This point was chosen as a reference as it was seen in all samples and at the same time (within slight variations of chromatographic performance). To further check the assignment of peaks, samples were mixed in known ratios and analysed.

This showed definitively that two peaks thought to be the same in two different samples were in fact the same component.

Mass spectral analysis of the samples was shown to provide definitive identification of the major by-products of synthesis. In conjunction with chromatographic retention time, peaks present at lower levels and hence with low quality spectra can still be identified in the samples. Limitations of sensitivity prevented the identification of all the peaks present in the samples of HFC 134a studied, especially those peaks seen in the ECD chromatographic traces.

8.3 Classification patterns of known and unknown samples

The data obtained from using the different detection systems produced a series of chromatographs where each peak in the chromatograph represented the presence of a by-product of production. The peak height values were measured and used as the value to describe the amount of component present. The data on which the characterisation was based consisted of a list of peak height values for each component present. If the component was not detected in the chromatographic trace an arbitrary value of half the value for the limit of detection was applied.

The chromatographic data was formed into matrices where the different samples analysed were represented by different rows and the columns represented the different components present. The values in each cell represented the peak heights for each component in each sample of HFC 134a. These matrices were subjected to the range of chemometric analyses studied in this work.

The samples analysed were split into three separate classes. The first class of samples were those from known sources of production and used initially to investigate the use of a number of chemometric techniques. The second class of samples were those from known sources or mixtures of samples from known sources which were used to test the classification protocols developed. The third class of samples were those of either unknown origin to the analyst but known to a third party

or of a totally unknown origin. The third set of samples were used to test the application of the chemometric techniques without the potential for operator interpretation *i.e.* the identity of the samples was only revealed after analysis and classification. The samples of totally unknown origin were analysed to check their identity and to see how the classification of potentially new samples (*i.e.* from a production origin not previously tested) would affect the chemometric techniques employed.

The analysis of the known samples by FID and ECD produced chromatographic profiles where a large number of synthesis by-products (>30) were separated from the parent peak and from each other. The traces gave vastly different profiles, in that the CFC and HCFC by-products of synthesis had large peak heights in the ECD traces in comparison with those seen in the FID traces (relative response for CFC 12). Also, peaks corresponding to HFC by-products gave vastly reduced responses in the ECD trace (relative response for HFC 134a). The chromatographic profiles formed by mass spectral analysis appeared most similar in height responses to those seen in the FID as the MS has no particular increased level of sensitivity to any one type of by-product.

The classification protocol to be followed to identify the origin of production of a sample of HFC 134a is displayed in Figure 8.1. The flow diagram begins with the acquisition of the sample. The analysis can then proceed to use either GC FID or GC ECD to chromatographically profile the by-products of synthesis. Analysis of a sample of HFC 134a by GC ECD has been shown to provide more peaks (because of higher levels of sensitivity) and hence more information to classify the peaks. However, either route will allow the correct classification of an unknown sample. The lines drawn in blue in the flow diagram show the minimum requirement for the classification of an unknown sample. That is, chromatographic profiling followed by PCA and hierarchical cluster analysis of the PC scores will allow the identification of a sample. However, the lines marked in red highlight extra analyses which will increase the level of confidence applied to the classification of an unknown sample.

The mass spectral analysis of an unknown sample will allow the definitive identification of by-products and hence provide higher levels of confidence to the alignment of peaks between chromatographic traces. Mixing an unknown with known samples will identify peaks between traces, although this is time consuming when trying to identify a number of peaks seen in different known samples. The further analysis of a sample which has not been closely classified to any samples of known origin provides a clearer picture of whether the samples are of the same origin of production.

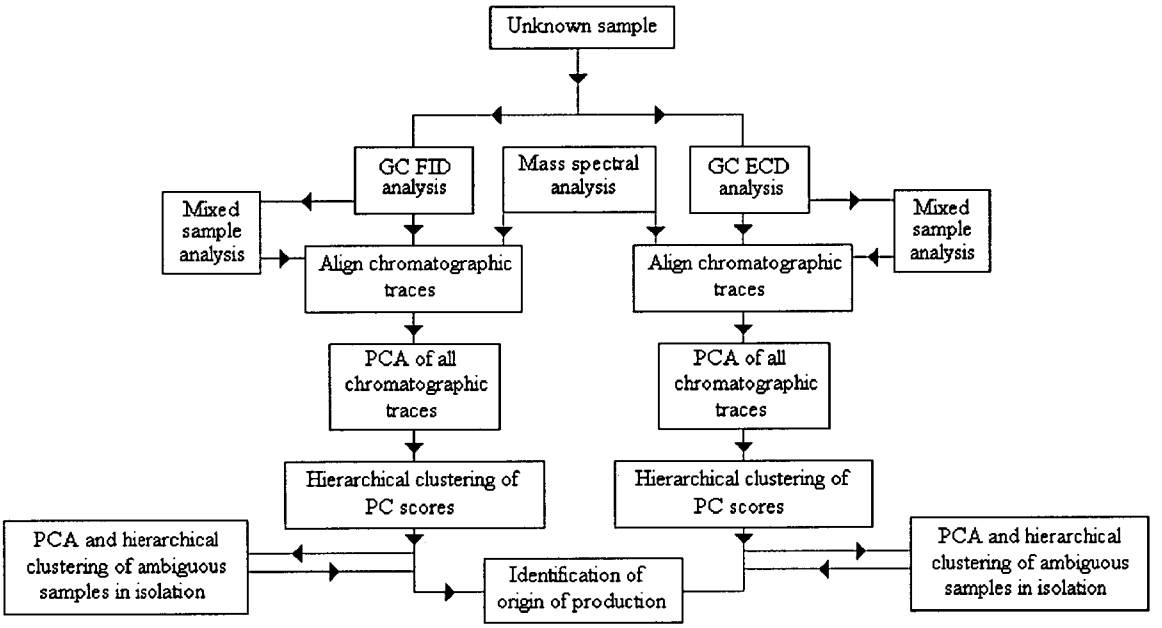


Figure 8.1 *Flow diagram of the classification protocol for an unknown sample of HFC 134a.*

Whichever analysis route is applied to initially identify an unknown sample it is obviously possible to repeat the classification of the sample using the other route of analysis. This will provide two independent results to confirm the identity of the sample of unknown origin.

The classification of samples performed in this work has correctly separated a number of different samples into their correct origins of production. The training

samples used to formulate the analysis protocol were all correctly placed into clusters on the basis of their origins of production.

The correct classification of test samples showed that the analysis method could identify samples not previously considered by the analysis method. The classification of the mixtures of samples placed these samples into areas in-between the samples of their composition, as would be expected. The exact distance from each cluster they were composed of related to the ratio of their composition.

Testing of the analysis methods by classifying samples which had their identities withheld until after they had been classified removed the possibility of analyst interpretation effecting the results. The analysis of samples whose origin of production was not known and could not be checked served to further test the analysis method and showed a number of samples that were not produced by any of the production sites investigated in this work. The continued analysis of samples from both originally tested and new sites of production would show whether these samples were from a different origin of production or were from known origins, but displaying large variations from the training and test samples.

8.4 Future considerations and progressions

The future progression of this work is the continued analysis of samples of HFC 134a. As with any statistical analysis technique, the increase in *n* number leads to an increase in the levels of confidence which can be applied to any results[13]. With the analysis of more samples of HFC 134a the method becomes more robust and the lines of classification between different samples become more clearly defined. The analysis of new samples from origins of production not considered in this work would also increase the scope of the answers to be gained *i.e.* the identity of the unknown samples seen in the sample set CE obtained from the Chemical Engineering Department may be discovered.

The routine analysis of samples obtained from a manufacturing process is one of the applications this work was designed to fulfil. The application of the techniques described would lead to the further testing of the analysis protocol. The *in situ* use of the GC analysis methods followed by the chemometric classification of samples would further show that even small changes within a samples composition of synthesis by-products affects the way in which it is classified[14]. The application of the analysis protocol in this way would require a certain amount of refinement as the boundaries between what is classed as an acceptable level of dissimilarity and what is classed as a 'good' sample have not been defined in this work. Therefore in the practical application of the analysis methods, samples would be analysed and monitored at a production plant. A set distance measure from a cluster would then be defined either on what was a straight assessment of purity or from later feedback from the end user as to the acceptable level of performance of a specific batch *i.e.* if a sample of HFC 134a was shown to have unacceptable qualities the position it occupied in multi-dimensional space would be flagged and any further samples of HFC 134a appearing close to those unacceptable areas in space would be rejected.

The further improvement of the software used to handle the data is also an identified progression. The ideal piece of software for the work performed here would be a program which controlled the analytical instrument, acquired the analytical data and performed the chemometric analysis. With the increase in the application and understanding of chemometric processes the closer relationship between the acquisition software and analysis software will surely occur. There seems little point in transferring data from one format on a computer to another format on a computer, if compatible, merely to perform a series of computationally simple calculations. The closer integration of chemometric analysis software will follow the same path as the integration of instrument control and data acquisition software to be performed by a single piece of software. This will allow the operator of a piece of analytical equipment to perform all the stages of an analysis to obtain a final result.

The chemometric techniques used in the analysis and classification of chromatographic data described in this work can easily be applied to other types of multi-variate data[15]. As discussed in the introduction it is not so much the specific analytical technique which dictates the use of a chemometric process but the form of data obtained. Any chemical process which produces results of a multi variate nature which need to be classified into groups on the basis of similarity/dissimilarity can have these methods applied. If this project had used the trace composition of metals found in the samples (either present in the raw materials or from catalysts in the synthesis methods) the results could have been treated in an identical manner. The only difference would be that the variables would now be the levels of metals in each sample as opposed to the levels of by-products present in the samples.

One specific area of research related closely to the work performed here would be the application of a similar protocol to the analysis of samples of CFCs. These compounds are currently still in use and production in a number of developing countries which have been granted longer time scales for CFC phase out. This has led to a huge market in both illegally produced and distributed CFCs[16]. Nowhere is this more apparent than across the Mexican/American border.

America has phased out the production of CFCs and is presently replacing them with either alternatives or reclaimed CFCs. For instance, when the air conditioning unit on a car runs low on CFC levels these either have to be 'topped up' with reclaimed CFCs or the system must be altered to be compatible with the new HFCs, both represent costly alternatives. However, as Mexico is still permitted to produce CFCs, it is possible to smuggle the CFCs into America to be sold illegally. These illegally sold CFCs can then be used to 'top up' the car's air conditioning unit at a lower price than would legally be possible. Presently, the chemical differentiation between legally reclaimed and smuggled CFCs is not possible. By applying the techniques described in this work it may be possible to identify batches of samples from different origins, and therefore identify those samples of smuggled CFCs from legal reclaimed CFCs. Although this may appear a small problem a report claimed

that after cocaine, illegally produced CFCs were the second most smuggled item to travel through Miami[17].

References

[1] McCulloch, A. and Campbell, N., Chemistry and Industry, **7**, p262-267, (1999)

[2] Chemistry in Britain, **October**, p10, (1997)

[3] Banks, R.E., Smart B.E. and Tatlow J.C., Organofluorine Chemistry Principles and Commercial Applications, 159-175 Plenum Press, New York and London

[4] de Zeeuw, J., de Nijs, R.C.M. and Henrich, L.T., J. Chromatogr. Sci., **25**, p71, (1987)

[5] Bachmann, K. and Reineke, F.J., Journal of Chromatography, **323**, p323-329, (1985)

[6] Lavine, B.K., Anal. Chem., **70**, 209R-228R, (1998)

[7] Rood, D., A Practical Guide to the Care, Maintenance and Troubleshooting of Capillary Gas Chromatographic systems, Huthig, Heidleberg, (1991)

[8] de Zeeuw, J, Zweip, D. and Marinissen, J.W., International Laboratory, **September**, 12J-12P, (1996)

[9] Sturrock, G.A., Simmonds, P.G., Nickless, G. and Zwiep, D., Journal of Chromatography, **648**, p423-431, (1993)

[10] MacDougall (Chairman) *et al.*, Anal. Chem., **52**, 2242, (1980)

[11] Danielsson, R. and Malmquist, G., Journal of Chromatography A, **687**, p71-78, (1994)

[12] Fowlis, I.A., Gas Chromatography, 2nd ed., John Wiley and Sons, Chichester, England, p179, (1995)

[13] Mead, R., Curnow, R.N. and Hasted, A.M., Statistical Methods in Agriculture and Experimental Biology, 2nd ed., Chapman and Hall, London, p393, (1993)

[14] Wilcken, H. and Schulten, H-R., Fresenius J. Anal. Chem., **335**, p157-163, (1996)

[15] Choo, L.P., Mansfield, J.R., Pizzi, N., Somorjai, R.L., Jackson, M., Halliday, W.C. and Manstech, H.H., Biospectroscopy, **1(2)**, p141-148, (1995)

[16] Spurgeon, D., Nature, **389**, 219 (1997)

[17] Chemistry and Industry, **9**, 315, (1996)

Appendix

Appendix A

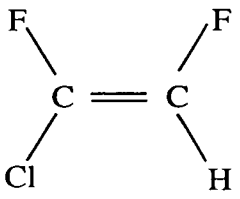
Hydrohalocarbons are identified using a shorthand representation made up of letters and numbers. The shorthand version describes to what class the hydrohalocarbon belongs, the molecule formula of the compound and the isomeric form. The shorthand form is made up of a series of letters which describe the series the compound belongs to:

CFC	=	ChloroFluoroCarbon.
HCFC	=	HydroChloroFluoroCarbon.
HFC	=	HydroFluoroCarbon.

The letter code is followed by a series of numbers

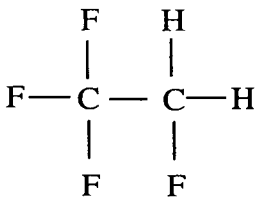
First digit	=	Number of Carbon double bonds.
Second digit	=	Number of Carbon atoms minus one.
Third digit	=	Number of Hydrogen atoms plus one.
Fourth digit	=	Number of Fluorine atoms.

The remaining atoms not accounted for in the numeric description are made up of Chlorine atoms. The isomeric form of the compound is identified by the lower case letter following after the numbers. The most symmetrical form of the compound has no letter, the second most symmetrical form has the letter a and further isomers are described with the next letter in alphabetical order. Isomeric forms of compounds containing more than 2 carbons are described by firstly assigning a letter to the central carbon in the chain followed by the carbons on either side of the central carbon. A series of examples is given below.



HCFC 1122

HCFC describes the compound as a HydroChloroFluoroCarbon. The first digit indicates a double bond. The second digit indicates two carbon atoms. The third digit indicates one hydrogen atom. The fourth digit indicates two fluorine atoms. The remaining unassigned bond is hydrogen. The absence of a lower case letter indicates the molecule is the most symmetrical isomer



HFC 134a

HFC describes the compound as a HydroFluoroCarbon. The first digit indicates two carbon atoms. The second digit indicates two hydrogen atoms. The third digit indicates four fluorine atoms. The lower case 'a' indicates the isomer is not the most symmetrical form.

List of components seen in HFC 134a from all manufacturing sources (International Union of Pure and Applied Chemistry)

Arcton	Code number	Expected in ICI produced HFC 134a
1,1-Dichloro-1,2,2,2-tetrafluoroethane	114a	Yes
Pentafluoroethane	125	Yes
1,1,1-Trifluoroethane	143a	Yes
Bromochlorodifluoromethane	12B1	Yes
2-Chloro-1,1,1-trifluoroethane	133a	Yes
Chloropentafluoroethane	115	Yes
Dichlorodifluoromethane	12	Yes
1,2-Dichloro-1,1,2,2-tetrafluoroethane	114	Yes
1,1,2,2-Tetrafluoroethane	134	Yes
Trichlorofluoromethane	11	No
Chlorodifluoromethane	22	Yes
1,1-Dichloro-2,2,2-trifluoroethane	123	No
1,2-Dichloro-1,2,2-trifluoroethane	123a	No
1,2-Dichloro-1,1,2,2-tetrafluoroethane	124a	No
1,1-Difluoroethane	152a	No
1,1,1,2,2-Pentafluoropropane	245cb	Yes
Methyl chloride	40	Yes
1,2-Difluoroethane	152	No
1,2-Dichloro-2,2-difluoroethane	132b	No
Chlorofluoromethane	31	No
2-Chloro-1,1,1,2-tetrafluoroethane	124	Yes
3,3,3-Trifluoroprop-1-ene	1243zf	Yes
2-Chloro-1,1-diofluoroethane	1122	No
1 Chloro-1,2-difluoroethene	1122a	Yes
2,3,3,3-Tetrafluoropropene	1234yf	Yes
1,1,3,3,3-Pentafluoroprop-1-ene	1225ye	No
Perfluorobut-2-ene	1318my	No
1-Chloro-2-fluoroethene	1131	No
1,1-Dichloro-2,2-difluoroethene	1112	No

Appendix C

Table detailing dilution volumes and steps to produce range of concentrations (nmoles)

250µl loop injections					**100µl loop injections**				
	1ml spiked	5ml spiked	10ml spiked	20ml spiked		1ml spiked	5ml spiked	10ml spiked	20ml spiked
Dilution 1	6.397E+0	3.191E+1	6.365E+1	1.266E+2	Dilution 1	2.559E+0	1.276E+1	2.546E+1	5.062E+1
Dilution 2	2.934E-2	1.461E-1	2.909E-1	5.766E-1	Dilution 2	1.173E-2	5.845E-2	1.164E-1	2.306E-1
Dilution 3	1.354E-4	6.747E-4	1.343E-3	2.661E-3	Dilution 3	5.417E-5	2.699E-4	5.372E-4	1.064E-3
Dilution 4	6.224E-7	3.101E-6	6.173E-6	1.223E-5	Dilution 4	2.489E-7	1.240E-6	2.469E-6	4.892E-6
50µl loop injections					**10µl loop injections**				
	1ml spiked	5ml spiked	10ml spiked	20ml spiked		1ml spiked	5ml spiked	10ml spiked	20ml spiked
Dilution 1	1.279E+0	6.382E+0	1.273E+1	2.531E+1	Dilution 1	2.556E-1	1.275E+0	2.543E+0	5.058E+0
Dilution 2	5.867E-3	2.923E-2	5.818E-2	1.153E-1	Dilution 2	1.172E-3	5.840E-3	1.163E-2	2.304E-2
Dilution 3	2.709E-5	1.349E-4	2.686E-4	5.322E-4	Dilution 3	5.412E-5	2.696E-5	5.367E-5	1.063E-4
Dilution 4	1.245E-7	6.201E-7	1.234E-6	2.446E-6	Dilution 4	2.487E-8	1.239E-7	2.467E-7	4.887E-7

APPENDIX D

Retention times	Jun-96	Aug-96	Oct-96	Dec-96	Jan-97	Feb-97	Mar-97	May-97	Aug-97	Oct-97	Jan-98	Feb-98	Jun-98	Jul-98
Methane	3.22	3.30	3.39	3.33	3.39	3.27	3.22	3.26	3.31	3.26	3.22	3.23	3.22	3.30
Ethane	3.60	3.70	3.79	3.80	3.80	3.66	3.60	3.65	3.70	3.65	3.61	3.61	3.61	3.70
Ethylene	4.00	4.10	4.21	4.30	4.21	4.07	4.00	4.05	4.11	4.05	4.01	4.01	4.01	4.11
Propane	5.27	5.46	5.65	5.62	5.66	5.34	5.27	5.32	5.37	5.32	5.27	5.27	5.27	5.47
Propylene	6.34	6.51	6.68	6.73	6.68	6.45	6.34	6.43	6.52	6.43	6.35	6.36	6.35	6.52
n-Butane	8.28	8.50	8.72	8.80	8.73	8.43	8.28	8.40	8.52	8.40	8.30	8.31	8.30	8.51
Acetylene	11.12	11.42	11.71	11.77	11.72	11.31	11.12	11.28	11.43	11.28	11.14	11.16	11.14	11.43
Propyne	13.10	13.45	13.80	13.64	13.81	13.32	13.10	13.28	13.47	13.28	13.12	13.14	13.12	13.46
Base width	Jun-96	Aug-96	Oct-96	Dec-96	Jan-97	Feb-97	Mar-97	May-97	Aug-97	Oct-97	Jan-98	Feb-98	Jun-98	Jul-98
Methane	0.119	0.104	0.110	0.119	0.108	0.107	0.114	0.107	0.110	0.107	0.105	0.101	0.109	0.110
Ethane	0.081	0.094	0.088	0.087	0.093	0.081	0.082	0.087	0.087	0.088	0.083	0.095	0.089	0.090
Ethylene	0.096	0.096	0.098	0.096	0.082	0.084	0.094	0.098	0.097	0.097	0.097	0.081	0.095	0.097
Propane	0.102	0.099	0.094	0.103	0.101	0.090	0.105	0.094	0.102	0.099	0.089	0.092	0.102	0.096
Propylene	0.097	0.102	0.089	0.092	0.102	0.103	0.100	0.091	0.096	0.098	0.087	0.098	0.093	0.101
n-Butane	0.139	0.131	0.134	0.138	0.134	0.144	0.143	0.145	0.138	0.139	0.138	0.137	0.139	0.132
Acetylene	0.194	0.191	0.188	0.190	0.188	0.193	0.198	0.197	0.200	0.181	0.197	0.195	0.185	0.199
Propyne	0.207	0.224	0.207	0.223	0.212	0.216	0.221	0.209	0.210	0.217	0.218	0.219	0.218	0.217
Resolution values	Jun-96	Aug-96	Oct-96	Dec-96	Jan-97	Feb-97	Mar-97	May-97	Aug-97	Oct-97	Jan-98	Feb-98	Jun-98	Jul-98
Methane-Ethane	3.85	3.98	4.10	4.05	4.04	4.17	3.92	4.03	4.01	3.99	4.10	3.95	3.89	3.97
Ethane-Ethylene	4.51	4.30	4.52	5.38	4.79	4.91	4.52	4.37	4.44	4.36	4.44	4.55	4.33	4.38
Ethylene-Propane	12.83	13.87	14.97	4.95	15.70	14.68	12.76	13.18	12.65	12.93	13.56	14.49	12.78	14.12
Propane-Propylene	10.82	10.49	11.27	13.69	10.12	11.44	10.50	12.06	11.57	11.29	12.31	11.48	11.13	10.59
Propylene-n-Butane	16.44	17.12	18.30	16.21	17.38	15.94	15.98	16.69	17.09	16.62	17.28	16.57	16.77	17.11
n-Butane-Acetylene	17.04	18.07	18.57	12.62	18.57	17.10	16.62	16.81	17.28	17.98	16.95	17.14	17.54	17.59
Acetylene-Propyne	9.88	9.78	10.58	14.34	10.43	9.84	9.44	9.90	9.94	10.08	9.54	9.58	9.84	9.76

Retention times	Nov-97	Dec-97	Feb-98	Mar-98	May-98
Methane	3.39	3.30	3.26	3.30	3.26
Ethane	3.80	3.70	3.65	3.69	3.65
Ethylene	4.22	4.11	4.05	4.10	4.05
Propane	5.68	5.47	5.31	5.35	5.31
Propylene	6.69	6.52	6.42	6.50	6.42
n-Butane	8.75	8.51	8.39	8.49	8.39
Acetylene	11.74	11.43	11.26	11.40	11.26
Propyne	13.83	13.46	13.26	13.43	13.26
Base width	Nov-97	Dec-97	Feb-98	Mar-98	May-98
Methane	0.115	0.112	0.101	0.100	0.108
Ethane	0.100	0.082	0.098	0.100	0.090
Ethylene	0.083	0.091	0.085	0.086	0.090
Propane	0.104	0.104	0.104	0.101	0.088
Propylene	0.103	0.096	0.101	0.103	0.098
n-Butane	0.143	0.135	0.145	0.138	0.131
Acetylene	0.194	0.197	0.180	0.192	0.189
Propyne	0.220	0.223	0.208	0.222	0.221
Resolution values	Nov-97	Dec-97	Feb-98	Mar-98	May-98
Methane-Ethane	3.78	4.08	3.91	3.95	3.93
Ethane-Ethylene	4.60	4.72	4.40	4.39	4.48
Ethylene-Propane	15.61	13.95	13.30	13.37	14.20
Propane-Propylene	9.78	10.41	10.85	11.27	11.97
Propylene-n-Butane	16.68	17.28	15.98	16.49	17.14
n-Butane-Acetylene	17.76	17.59	17.64	17.59	17.93
Acetylene-Propyne	10.10	9.70	10.32	9.79	9.78

Appendix E

Stability monitoring of FID signal over time course of work.

Date analysis performed		Response (area units)	Date analysis performed		Response (area units)
July 96	2	379373721	September 97	17	302893120
		366274240			314157777
		365886651			341682560
		365136320		22	322652830
	8	367174870			322652830
		373736229	November 97	12	313859200
		371611283		13	331822192
	15	336724916			331620386
		336616000			342661120
		336506560	January 98	13	316835520
	23	352038812			312408353
		352038812			311163840
		355602271		26	347302995
	30	335067200			347302995
		334527135			345327019
		334328960	February 98	10	334620800
October 96	4	359719396		12	368458542
		357238718			367352320
		356200961	March 98	20	317847692
		353797429			317224900
November 96		343538560		22	336361984
		342670400			336064000
		342670400			335734272
December 96	2	302799166			335734016
		302354649	May 98	03	323457280
		301480960			323457280
January 97	07	332422156	June 98	07	326563200
		332181120		14	345327019
		332101099			343538560
	17	309488168		20	339449920
February 97	12	309033909			338613843
March 97	13	324344699			338604160
		324344699			342661120
		323457280			334045365
	25	335127040	July 98	03	329953280
April 97	30	342428928			329953280
		342428928			331149353
Jun 97	27	364549760			341682560
		363870720		10	341878009
		361115200			

Appendix F

Stability monitoring of ECD signal over time course of work.

November 96	01	364105728	October 97	07	339885760
December 96	11	278592000			380640000
		280500000		16	384521920
January 97	16	359393280			319512480
		336494080	November 97	12	335519520
		373986000			406598731
	19	379283440			345698564
		365543040		20	389872320
February 97	19	310345120	January 98	13	392640000
March 97	13	362516874		16	381095360
	15	366686720		29	374364800
		373302720			349494880
		365246987	March 98	1	395277760
		342079680			358540960
April 97	02	395479520		11	383008000
		376977920		14	391067360
	05	377008000			377952000
		327808000			369408000
		361952000	April 98	20	374482080
July 97	05	363660320			402698712
		356922720			362720000
	13	387545120	May 98	4	358688000
		361292960			353668701
		342659873		11	342421926
August 97	20	345987129			369600000
		375808000		18	303536000
		364587951			340357795
September 97	03	317324160	June 98	1	369600000
		370816000			340256000
	20	391316960		9	358320000
		353770720			358464000
		326598774			371408000
				13	347375726

Appendix G

FID replicate analyses of HFC 134a produced by ICI UK

Replicate 1			Replicate 2			Replicate 3			Replicate 4			Replicate 5		
Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height
2.63	0.089	0.618	2.57	0.087	0.540	2.65	0.084	0.665	2.66	0.084	0.752	2.65	0.085	0.680
13.02	0.440	3.723	13.00	0.441	3.791	13.25	0.420	3.020	13.27	0.418	3.077	13.25	0.424	2.947
22.11	0.746	0.518	21.98	0.745	0.553	22.63	0.718	0.455	22.67	0.714	0.467	22.57	0.722	0.420
23.23	0.784	0.085	23.04	0.781	0.096	23.74	0.753	0.081	23.81	0.749	0.071	23.70	0.758	0.073
25.83	0.872	0.081	25.84	0.876	0.072	26.68	0.846	0.051	26.77	0.843	0.061	26.59	0.851	0.053
26.62	0.899	2.154	26.50	0.899	2.134	27.21	0.863	1.220	27.85	0.877	1.717	27.54	0.881	1.599
27.52	0.929	3.969	27.43	0.930	3.865	28.06	0.890	2.876	29.01	0.913	2.897	28.66	0.917	2.593
28.63	0.967	1.161	28.56	0.968	0.865	30.16	0.957	0.651	30.39	0.957	0.681	29.99	0.959	0.617
29.51	0.996	0.607	29.39	0.997	0.601	31.27	0.992	0.471	31.54	0.993	0.489	31.07	0.994	0.465
29.62	*1.000*	*1.000*	*29.49*	*1.000*	*1.000*	*31.52*	*1.000*	*1.000*	*31.77*	*1.000*	*1.000*	*31.26*	*1.000*	*1.000*
47.96	1.619	7.570	54.43	1.846	8.077	49.07	1.557	7.016	49.12	1.546	7.636	49.43	1.581	6.935
49.33	1.665	3.048	56.19	1.905	1.847	50.41	1.599	2.529	50.46	1.588	2.692	50.49	1.615	2.417
56.68	1.914	0.356	63.21	2.143	0.519	57.72	1.831	0.388	57.75	1.818	0.397	57.73	1.847	0.322
62.75	2.119	14.447	69.52	2.357	5.692	64.42	2.044	16.172	64.62	2.034	8.775	64.68	2.069	13.335
68.75	2.321	4.768	74.96	2.542	1.885	71.84	2.279	4.825	70.29	2.212	2.208	70.33	2.250	4.395
69.15	2.335	1.672	76.30	2.587	2.308	72.28	2.293	1.828	71.66	2.256	1.311	71.73	2.295	1.606

Rt = retention time in minutes
RRt = relative retention time (to leading edge for HFC 134a peak)
Height = height response in mV at sensitivity setting R0

FID replicate analyses of HFC 134a produced by ICI USA.

Replicate 1			Replicate 2			Replicate 3			Replicate 4			Replicate 5		
Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height
2.650	0.089	0.505	2.680	0.088	0.532	2.670	0.089	0.471	2.680	0.086	0.619	2.660	0.087	0.569
3.069	0.103	1.079	3.070	0.101	1.043	3.060	0.102	1.062	3.060	0.098	0.942	3.060	0.100	1.046
4.200	0.141	8.103	4.230	0.140	7.833	4.220	0.141	8.022	4.220	0.136	7.183	4.230	0.139	7.826
4.820	0.162	0.163	4.840	0.160	0.159	4.840	0.161	0.159	4.860	0.156	0.155	4.860	0.159	0.180
12.770	0.428	5.237	1.282	0.423	4.617	12.850	0.428	4.908	12.950	0.417	4.197	12.920	0.424	4.738
12.860	0.431	36.092	12.910	0.426	35.058	12.930	0.431	36.283	13.050	0.420	33.034	13.010	0.427	34.558
13.000	0.436	43.669	13.060	0.431	41.652	13.080	0.436	43.854	13.200	0.425	37.907	13.197	0.433	38.349
14.360	0.481	0.113	14.440	0.477	0.117	14.470	0.482	0.175	14.600	0.470	0.113	14.540	0.477	0.102
17.400	0.583	0.505	17.470	0.577	0.486	17.490	0.582	0.508	17.700	0.569	0.438	17.600	0.577	0.494
17.930	0.601	0.084	17.960	0.593	0.082	17.980	0.599	0.081	18.200	0.586	0.066	18.100	0.593	0.072
19.220	0.644	0.073	19.300	0.637	0.058	19.180	0.639	0.061	19.500	0.628	0.061	19.450	0.638	0.060
20.950	0.702	0.051	21.050	0.695	0.067	21.040	0.701	0.060	21.310	0.686	0.045	21.210	0.695	0.050
25.800	0.865	2.463	26.160	0.863	2.348	25.970	0.865	2.476	26.470	0.952	2.143	26.210	0.859	2.352
28.840	0.967	0.141	29.190	0.963	0.141	29.040	0.967	0.139	29.810	0.959	0.139	29.410	0.964	0.141
29.830	*1.000*	*1.000*	*30.300*	*1.000*	*1.000*	*30.030*	*1.000*	*1.000*	*31.070*	*1.000*	*1.000*	*30.500*	*1.000*	*1.000*
-	-	-	51.070	1.685	3.715	-	-	-	48.850	1.572	2.019	50.490	1.655	2.777
50.740	1.701	0.457	-	-	-	51.040	1.700	0.575	51.040	1.643	0.576	-	-	-
-	-	-	61.737	2.038	1.090	-	-	-	61.353	1.975	4.662	62.570	2.051	4.154
65.620	2.200	2.432	65.660	2.167	4.167	65.680	2.187	4.231	65.740	2.116	4.774	65.760	2.156	4.615
-	-	-	-	-	-	67.670	2.253	1.704	67.820	2.183	1.473	67.910	2.227	1.661
71.220	2.388	1.569	71.290	2.353	1.809	71.330	2.375	2.215	71.410	2.298	2.053	71.425	2.342	2.000

Rt = retention time in minutes, RRt = relative retention time (to leading edge for HFC 134a peak), Height = height response in mV at sensitivity setting R0

FID replicate analyses of HFC 134a produced by ICI Japan.

Replicate 1			Replicate 2			Replicate 3			Replicate 4			Replicate 5		
Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height
2.660	0.084	0.567	2.570	0.084	0.451	2.660	0.086	0.508	2.680	0.850	0.595	2.660	0.084	0.824
24.100	0.764	0.241	23.630	0.771	0.253	23.870	0.769	0.242	24.050	0.767	0.242	24.170	0.759	0.227
28.040	0.889	0.212	27.430	0.896	0.227	27.720	0.892	0.202	27.960	0.891	0.201	28.230	0.887	0.190
30.290	0.960	0.283	29.500	0.963	0.294	29.860	0.961	0.262	30.120	0.960	0.282	30.470	0.957	0.235
31.540	*1.000*	*1.000*	*30.630*	*1.000*	*1.000*	*31.060*	*1.000*	*1.000*	*31.370*	*1.000*	*1.000*	*31.830*	*1.000*	*1.000*
50.650	1.606	0.711	50.490	1.661	0.780	50.600	1.629	0.681	50.680	1.616	0.687	50.700	1.593	0.669
54.180	1.719	2.691	-	-	-	55.250	1.779	2.353	-	-	-	-	-	-
65.730	2.084	5.639	59.160	1.931	5.287	65.730	2.116	5.273	65.755	2.096	4.754	65.757	2.066	4.664
66.790	2.118	0.135	-	-	-	66.790	2.150	0.199	66.830	2.130	0.094	66.830	2.099	0.117
68.400	2.169	0.245	-	-	-	-	-	-	-	-	-	68.540	2.153	0.273
71.420	2.264	2.183	65.728	2.146	1.770	71.410	2.300	1.893	71.460	2.278	1.910	71.400	2.243	1.908
73.130	2.320	1.659	-	-		73.130	2.354	1.610	73.250	2.335	1.686	73.220	2.300	1.846

Rt = retention time in minutes
RRt = relative retention time (to leading edge for HFC 134a peak)
Height = height response in mV at sensitivity setting R0

FID replicate analyses of HFC 134a produced by Atochem.

Replicate 1			Replicate 2			Replicate 3			Replicate 4			Replicate 5		
Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height
2.700	0.092	0.484	2.680	0.090	0.500	2.650	0.088	0.484	2.680	0.089	0.582	2.980	0.087	0.782
13.100	0.448	0.095	13.160	0.443	0.104	13.230	0.437	0.092	13.250	0.439	0.091	13.360	0.434	0.093
21.860	0.748	0.981	22.070	0.743	0.915	22.300	0.737	0.858	22.330	0.740	0.862	22.560	0.734	0.789
22.850	0.782	0.817	23.110	0.778	0.793	23.310	0.770	0.713	23.360	0.774	0.692	23.640	0.769	0.665
26.290	0.900	1.222	26.630	0.896	1.139	26.980	0.891	1.052	26.960	0.893	1.066	27.360	0.889	0.988
27.250	0.933	0.863	27.630	0.930	0.750	28.010	0.925	0.706	27.990	0.927	0.731	28.430	0.925	0.665
28.160	0.964	0.104	28.520	0.960	0.135	28.950	0.956	0.101	28.880	0.957	0.113	29.340	0.954	0.112
28.300	0.969	0.360	28.680	0.965	0.354	29.090	0.961	0.319	29.040	0.962	0.325	29.520	0.960	0.248
29.170	0.999	1.595	29.650	0.998	1.908	30.120	0.995	1.883	30.050	0.996	1.934	30.570	0.994	1.839
29.210	*1.000*	*1.000*	*29.710*	*1.000*	*1.000*	*30.270*	*1.000*	*1.000*	*30.180*	*1.000*	*1.000*	*30.750*	*1.000*	*1.000*
49.872	1.707	16.111	49.930	1.681	15.923	49.960	1.650	15.099	50.060	1.657	15.185	50.000	1.626	15.305
51.340	1.758	1.151	51.390	1.730	0.824	51.450	1.700	0.915	51.470	1.705	0.881	51.490	1.674	0.920
58.730	2.011	0.623	58.780	1.978	0.622	58.780	1.942	0.183	58.830	1.949	0.500	58.830	1.913	0.497
65.708	2.249	1.039	65.720	2.212	2.375	65.700	2.170	0.560	65.732	2.178	3.442	65.732	2.138	2.659
70.860	2.426	1.065	71.330	2.401	1.743	71.340	2.357	2.197	71.370	2.365	2.062	71.370	2.321	1.799

Rt = retention time in minutes
RRt = relative retention time (to leading edge for HFC 134a peak)
Height = height response in mV at sensitivity setting R0

FID replicate analyses of HFC 134a produced by Hoechst.

Replicate 1			Replicate 2			Replicate 3			Replicate 4			Replicate 5		
Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height
2.660	0.090	0.397	2.660	0.085	0.494	2.650	0.085	0.500	2.650	0.085	0.567	2.660	0.085	0.489
3.370	0.114	0.154	3.350	0.107	0.121	3.340	0.107	0.132	3.340	0.107	0.122	3.340	0.102	0.129
3.420	0.115	3.340	3.420	0.109	2.795	3.420	0.110	2.891	3.420	0.110	2.897	3.430	0.110	2.837
3.570	0.121	0.112	3.570	0.114	0.190	3.560	0.114	0.114	3.560	0.114	0.100	3.540	0.113	0.078
4.240	0.143	0.090	4.230	0.135	0.068	4.220	0.135	0.072	4.220	0.135	0.082	4.220	0.135	0.080
6.470	0.218	0.161	6.440	0.206	0.158	6.460	0.206	0.155	6.440	0.207	0.142	6.450	0.206	0.131
7.440	0.251	0.221	7.400	0.236	0.198	7.430	0.237	0.191	7.400	0.238	0.189	7.430	0.238	0.187
7.950	0.268	0.559	7.890	0.252	0.513	7.930	0.254	0.504	7.920	0.253	0.505	7.940	0.245	0.493
8.680	0.293	9.025	8.970	0.286	8.073	8.700	0.277	8.235	8.640	0.279	8.130	8.650	0.277	8.051
13.240	0.447	26.643	13.250	0.423	23.506	13.260	0.424	23.904	13.240	0.424	24.068	13.270	0.424	23.429
14.410	0.486	25.333	14.470	0.462	22.237	14.450	0.463	22.997	14.440	0.463	23.001	14.450	0.462	22.826
21.100	0.712	2.003	21.470	0.686	1.613	21.390	0.687	1.644	21.430	0.685	1.684	21.370	0.683	1.707
22.320	0.754	0.359	22.850	0.730	0.443	22.750	0.730	0.281	22.790	0.728	0.312	22.710	0.726	0.320
23.290	0.790	5.541	23.880	0.762	4.616	23.810	0.764	4.598	23.850	0.762	4.490	23.770	0.760	4.629
25.910	0.875	0.889	26.750	0.854	0.713	26.730	0.856	0.730	26.700	0.856	0.745	26.750	0.855	0.722
26.470	0.894	7.899	27.430	0.876	6.745	27.390	0.878	6.709	27.380	0.877	6.863	27.420	0.877	6.923
26.690	0.901	9.092	27.660	0.883	7.336	27.600	0.885	7.555	27.610	0.884	7.748	27.630	0.883	7.769
27.650	0.933	5.188	28.800	0.920	3.938	28.730	0.921	3.911	28.730	0.920	3.975	28.770	0.920	3.981
28.630	0.967	0.226	29.930	0.956	0.191	29.870	0.957	0.187	29.870	0.956	0.247	29.960	0.958	0.178
28.770	0.971	0.529	30.110	0.961	0.425	30.070	0.963	0.423	30.040	0.963	0.242	30.110	0.963	0.426
29.430	0.994	0.144	30.970	0.989	0.095	30.920	0.990	0.112	30.880	0.990	0.130	30.990	0.991	0.111
29.570	0.998	0.769	31.150	0.994	0.615	31.050	0.995	0.637	31.040	0.994	0.667	31.130	0.995	0.615

FID replicate analyses of HFC 134a produced by Hoechst continued.

Replicate 1			Replicate 2			Replicate 3			Replicate 4			Replicate 5		
Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height
29.620	*1.000*	*1.000*	*31.320*	*1.000*	*1.000*	*31.230*	*1.000*	*1.000*	*31.200*	*1.000*	*1.000*	*31.280*	*1.000*	*1.000*
51.200	1.729	10.549	51.190	1.634	7.844	51.230	1.628	7.621	51.380	1.645	7.993	51.400	1.643	7.969
56.750	1.916	2.902	-	-	-	52.520	1.683	2.269	-	-	-	52.710	1.685	1.685
58.590	1.988	21.194	58.600	1.871	18.947	58.650	1.880	18.609	58.690	1.879	19.537	58.680	1.876	19.961
63.112	2.131	3.028	-	-	-	61.110	1.959	6.523	-	-	-	61.380	1.962	5.885
65.590	2.214	1.798	65.570	2.094	2.858	65.610	2.103	2.268	65.645	2.102	2.314	65.660	2.099	2.692
71.218	2.404	1.514	71.210	2.274	1.800	71.280	2.285	1.660	71.300	2.283	1.667	71.330	2.280	1.981

Rt = retention time in minutes
RRt = relative retention time (to leading edge for HFC 134a peak)
Height = height response in mV at sensitivity setting R0

FID replicate analyses of HFC 134a produced by Showa Denka.

Replicate 1			Replicate 2			Replicate 3			Replicate 4			Replicate 5		
Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height
2.680	0.090	0.351	2.660	0.086	0.523	2.680	0.087	0.453	2.650	0.086	0.541	2.660	0.086	0.516
3.580	0.120	0.210	3.580	0.116	0.200	3.580	0.116	0.193	3.570	0.115	0.178	3.580	0.115	0.182
8.940	0.299	0.070	8.890	0.291	0.061	9.000	0.291	0.050	9.020	0.291	0.051	8.980	0.289	0.053
13.320	0.445	0.654	13.380	0.434	0.593	13.420	0.434	0.580	13.430	0.434	0.562	13.440	0.432	0.548
22.460	0.751	0.051	22.690	0.736	0.051	22.750	0.736	0.051	22.750	0.735	0.051	22.810	0.734	0.052
23.490	0.785	0.043	23.820	0.773	0.041	23.820	0.771	0.040	23.840	0.770	0.050	23.890	0.768	0.052
28.950	0.968	0.055	29.690	0.963	0.041	29.780	0.963	0.049	29.820	0.963	0.032	29.900	0.962	0.053
29.910	*1.000*	*1.000*	*30.620*	*1.000*	*1.000*	*30.910*	*1.000*	*1.000*	*30.970*	*1.000*	*1.000*	*31.090*	*1.000*	*1.000*
50.360	1.684	0.627	50.280	1.631	0.392	-	-	-	50.590	1.634	0.388	50.580	1.627	0.383
-	-	-	51.600	1.674	0.153	-	-	-	51.660	1.668	0.150	51.740	1.664	0.193
65.680	2.196	0.936	65.698	2.132	2.170	65.700	2.126	2.544	65.740	2.123	2.437	65.720	2.114	2.500
66.720	2.231	0.081	66.750	2.166	0.092	66.770	2.160	0.079	66.770	2.156	0.098	66.760	2.147	0.110
71.360	2.386	1.268	71.350	2.315	1.723	71.370	2.309	1.924	71.380	2.305	1.846	71.400	2.297	1.900

Rt = retention time in minutes
RRt = relative retention time (to leading edge for HFC 134a peak)
Height = height response in mV at sensitivity setting R0

FID replicate analyses of HFC 134a produced by Ashai Glass.

Replicate 1			Replicate 2			Replicate 3			Replicate 4			Replicate 5		
Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height
2.650	0.087	0.455	2.630	0.086	0.559	2.660	0.085	0.668	2.657	0.086	0.647	2.625	0.085	0.580
13.020	0.428	0.697	13.020	0.425	0.684	13.170	0.420	0.620	13.070	0.423	0.611	13.057	0.423	0.667
30.430	*1.000*	*1.000*	*30.670*	*1.000*	*1.000*	*31.330*	*1.000*	*1.000*	*30.898*	*1.000*	*1.000*	*30.868*	*1.000*	*1.000*
49.920	1.640	1.321	49.970	1.629	1.406	50.070	1.600	1.394	50.012	1.619	1.432	49.962	1.619	1.541
51.140	1.681	0.170	51.220	1.670	0.162	51.280	1.637	0.173	51.196	1.657	0.177	51.141	1.657	0.167
61.733	2.035	1.543	65.518	2.136	3.539	65.602	2.094	2.696	65.546	2.121	2.990	65.495	2.122	2.895
66.520	2.186	0.093	66.540	2.169	0.115	66.640	2.127	0.130	66.676	2.158	0.111	66.698	2.161	0.126
71.090	2.336	1.651	71.110	2.319	2.004	71.770	2.272	2.068	71.390	2.311	2.090	71.329	2.311	2.104
-	-	-	72.170	2.353	2.491	72.347	2.309	3.249	72.233	2.338	2.646	72.534	2.350	2.300

Rt = retention time in minutes
RRt = relative retention time (to leading edge for HFC 134a peak)
Height = height response in mV at sensitivity setting R0

Appendix H

ECD replicate analyses of HFC 134a produced by ICI in the UK.

Replicate 1			Replicate 2			Replicate 3			Replicate 4			Replicate 5		
Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height
2.619	0.082	127.6	2.631	0.083	112.9	2.615	0.083	116.6	2.629	0.083	116.0	2.615	0.084	105.1
6.292	0.197	0.124	6.308	0.199	0.176	6.269	0.199	0.132	6.273	0.198	0.160	6.288	0.202	0.120
6.420	0.201	0.964	6.435	0.203	1.044	6.395	0.203	0.836	6.368	0.201	0.968	6.382	0.205	1.048
8.496	0.266	0.184	8.464	0.267	0.192	8.442	0.268	0.164	8.395	0.265	0.172	8.405	0.270	0.164
8.656	0.271	4.308	8.622	0.272	3.880	8.600	0.273	2.348	8.554	0.270	2.444	8.561	0.275	2.640
21.304	0.667	76.88	21.334	0.673	81.20	21.231	0.674	75.91	21.226	0.670	76.37	21.075	0.677	85.15
26.830	0.840	0.460	26.755	0.844	0.508	26.681	0.847	0.448	26.770	0.845	0.452	26.616	0.855	0.484
27.628	0.865	469.7	27.547	0.869	503.6	27.437	0.871	479.3	27.562	0.870	486.2	27.083	0.870	527.6
28.267	0.885	1016	27.896	0.880	1063	27.783	0.882	1039	27.942	0.882	1034	27.768	0.892	1139
28.810	0.902	3440	28.689	0.905	3638	28.539	0.906	3516	28.702	0.906	3554	28.484	0.915	3877
31.876	0.998	1.228	31.605	0.997	2.616	-	-	-	-	-	-	-	-	-
31.940	*1.000*	*1.972*	*31.700*	*1.000*	*1.632*	*31.500*	*1.000*	*1.532*	*31.680*	*1.000*	*1.404*	*31.130*	*1.000*	*1.204*
38.775	1.214	25.89	38.484	1.214	15.02	38.525	1.223	3.396	38.364	1.211	2.732	38.383	1.233	4.628
49.411	1.547	1.416	49.928	1.575	1.416	48.825	1.550	1.476	49.009	1.547	1.292	49.248	1.582	1.400
50.817	1.591	23.46	50.657	1.598	20.86	50.211	1.594	24.00	50.435	1.592	20.01	50.680	1.628	22.15
57.875	1.812	28.92	58.360	1.841	4.384	57.929	1.839	4.212	58.164	1.836	4.444	58.306	1.873	4.708
58.354	1.827	3.320	60.737	1.916	1.248	60.386	1.917	0.812	60.572	1.912	0.908	60.641	1.948	1.108
-	-	-	63.974	2.018	7.100	-	-	-	-	-	-	63.882	2.052	0.264
64.679	2.025	6.276	64.700	2.041	11.51	64.418	2.045	2.508	64.500	2.036	2.772	64.595	2.075	4.280
65.573	2.053	791.5	65.492	2.066	621.1	65.300	2.073	239.6	65.356	2.063	217.5	65.435	2.102	356.0
65.892	2.063	3.076	65.873	2.078	3.088	65.615	2.083	3.100	65.704	2.074	3.924	65.778	2.113	3.876
66.595	2.085	195.2	66.507	2.098	192.4	66.213	2.102	179.9	66.338	2.094	166.1	66.431	2.134	214.4
69.406	2.173	13.66	69.328	2.187	11.88	69.080	2.193	10.56	69.189	2.184	10.10	69.295	2.226	14.21

ECD replicate analyses of HFC 134a produced by ICI in the USA.

Replicate 1			Replicate 2			Replicate 3			Replicate 4			Replicate 5		
Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height
2.616	0.080	119.7	2.629	0.081	113.0	2.617	0.080	116.1	2.610	0.080	116.2	2.616	0.078	135.9
3.107	0.095	158.0	3.116	0.096	167.1	3.107	0.095	164.0	3.099	0.095	158.5	3.086	0.092	136.6
3.368	0.103	0.292	3.376	0.104	0.284	3.369	0.103	0.280	3.392	0.104	0.292	3.388	0.101	0.240
5.821	0.178	0.120	5.648	0.174	0.084	5.822	0.178	0.080	5.839	0.179	0.084	5.870	0.175	0.124
6.377	0.195	1.124	6.395	0.197	1.128	6.378	0.195	1.084	6.394	0.196	1.076	6.406	0.191	0.948
6.475	0.198	0.196	6.492	0.200	0.332	6.477	0.198	0.296	6.524	0.200	0.324	6.540	0.195	0.304
8.731	0.267	16.74	8.732	0.269	17.01	8.766	0.268	16.95	8.807	0.270	16.50	8.821	0.263	15.35
13.015	0.398	16.06	13.049	0.402	16.85	13.084	0.400	16.45	13.113	0.402	15.80	13.215	0.394	14.32
16.121	0.493	0.164	16.100	0.496	0.204	16.224	0.496	0.220	16.245	0.498	0.248	16.435	0.490	0.172
17.854	0.546	0.164	17.853	0.550	0.164	17.925	0.548	0.168	18.006	0.552	0.168	18.179	0.542	0.208
21.615	0.661	13.50	21.618	0.666	14.15	21.687	0.663	14.00	21.758	0.667	13.44	22.036	0.657	11.89
26.879	0.822	0.200	26.812	0.826	0.216	26.953	0.824	0.200	27.042	0.829	0.204	27.436	0.818	0.296
27.272	0.834	22.55	27.201	0.838	22.65	27.313	0.835	22.75	27.368	0.839	23.09	27.805	0.829	19.98
28.318	0.866	2.120	28.208	0.869	1.956	28.327	0.866	1.760	28.412	0.871	1.400	28.945	0.863	0.756
28.613	0.875	2.252	28.532	0.879	2.104	28.654	0.876	2.020	28.706	0.880	1.808	29.247	0.872	1.280
29.561	0.904	14.16	29.441	0.907	13.19	29.570	0.904	11.920	29.619	0.908	9.808	30.220	0.901	5.348
32.700	*1.000*	*1.556*	*32.460*	*1.000*	*1.532*	*32.710*	*1.000*	*1.532*	*32.620*	*1.000*	*1.600*	*33.540*	*1.000*	*1.528*
39.632	1.212	18.412	39.699	1.223	18.31	39.645	1.212	20.00	39.731	1.218	22.94	39.846	1.188	27.85
51.993	1.590	147.42	49.274	1.518	145.6	49.327	1.508	146.7	49.419	1.515	153.7	49.237	1.468	178.6
62.032	1.897	1.100	62.064	1.912	1.032	62.051	1.897	1.104	62.076	1.903	0.976	62.015	1.849	1.196
63.405	1.939	0.196	64.855	1.998	0.168	64.897	1.984	0.268	64.914	1.990	0.212	63.726	1.900	0.440
65.858	2.014	2.976	65.861	2.029	2.916	65.878	2.014	3.172	65.892	2.020	3.128	65.839	1.963	3.648
66.708	2.040	329.4	66.705	2.055	338.1	66.761	2.041	362.7	66.741	2.046	431.2	66.678	1.988	541.9
67.100	2.052	4.604	67.095	2.067	4.024	67.121	2.052	4.468	67.132	2.058	5.144	67.046	1.999	3.692
67.853	2.075	165.0	67.744	2.087	163.8	67.873	2.075	177.1	67.784	2.078	187.5	67.717	2.019	211.8
70.665	2.161	8.800	70.633	2.176	8.684	70.654	2.160	8.920	70.655	2.166	11.63	70.568	2.104	12.38

ECD replicate analyses of HFC 134a produced by ICI in the Japan.

Replicate 1			Replicate 2			Replicate 3			Replicate 4			Replicate 5		
Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height
2.608	0.078	30.361	2.628	0.077	133.5	2.590	0.076	132.7	2.623	0.077	128.4	2.630	0.076	145.4
-	-	-	3.106	0.091	0.033	3.101	0.091	0.032	3.100	0.091	0.042	3.114	0.090	0.023
6.519	0.195	0.211	6.553	0.192	0.766	6.543	0.192	0.762	6.576	0.193	0.798	6.574	0.190	0.741
8.826	0.264	12.58	8.874	0.260	47.90	8.861	0.260	47.68	8.858	0.260	48.40	8.927	0.258	43.46
22.064	0.660	0.488	22.287	0.653	1.967	22.288	0.654	1.964	22.384	0.657	2.004	22.490	0.650	1.704
24.604	0.736	0.051	24.813	0.727	0.149	24.844	0.729	0.141	24.871	0.730	0.154	25.016	0.723	0.102
27.881	0.834	0.031	28.260	0.828	0.041	28.218	0.828	0.043	28.278	0.830	0.040	28.510	0.824	0.031
28.917	0.865	0.092	29.454	0.863	0.424	29.445	0.864	0.434	29.436	0.864	0.449	29.791	0.861	0.364
29.285	0.876	0.032	29.761	0.872	0.072	29.752	0.873	0.084	29.777	0.874	0.071	30.067	0.869	0.066
30.254	0.905	1.486	30.785	0.902	6.178	30.740	0.902	6.206	30.833	0.905	6.330	31.175	0.901	5.778
31.324	0.937	0.040	31.912	0.935	0.098	31.865	0.935	0.103	31.924	0.937	0.110	32.316	0.934	0.084
33.430	*1.000*	*0.450*	*34.130*	*1.000*	*1.522*	*34.080*	*1.000*	*1.458*	*34.070*	*1.000*	*1.470*	*34.600*	*1.000*	*2.100*
38.946	1.165	0.831	39.796	1.166	4.077	39.805	1.168	4.185	39.862	1.170	4.782	39.928	1.154	5.907
61.143	1.829	0.270	62.185	1.822	1.125	62.196	1.825	1.023	62.246	1.827	1.077	62.142	1.796	1.012
63.885	1.911	0.032	64.949	1.903	0.069	64.956	1.906	0.085	64.971	1.907	0.078	65.429	1.891	0.074
64.219	1.921	0.074	65.257	1.912	0.285	65.297	1.916	0.269	65.278	1.916	0.271	65.740	1.900	0.280
64.888	1.941	0.610	65.905	1.931	2.416	65.945	1.935	2.135	65.960	1.936	2.296	65.878	1.904	2.478
65.723	1.966	84.26	66.758	1.956	388.9	66.865	1.962	355.5	66.777	1.960	386.3	66.709	1.928	485.4
66.091	1.977	1.293	67.134	1.967	4.062	67.172	1.971	4.031	67.186	1.972	3.854	67.124	1.940	3.254
66.392	1.986	0.118	67.475	1.977	0.185	67.512	1.981	0.284	67.493	1.981	0.416	-	-	-
66.760	1.997	38.70	67.816	1.987	113.6	67.853	1.991	160.2	67.833	1.991	170.2	67.781	1.959	178.4
69.668	2.084	2.167	70.717	2.072	9.230	70.716	2.075	8.615	70.763	2.077	8.906	70.653	2.042	10.40

ECD replicate analyses of HFC 134a produced by Atochem.

Replicate 1			Replicate 2			Replicate 3			Replicate 4			Replicate 5		
Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height
2.620	0.081	27.21	2.636	0.081	79.52	2.615	0.081	84.90	2.622	0.082	87.45	2.616	0.082	81.54
6.535	0.202	0.159	6.541	0.201	0.253	6.553	0.203	0.114	6.556	0.205	0.247	6.540	0.205	0.285
8.832	0.273	1.072	8.851	0.272	1.296	8.845	0.274	1.423	8.858	0.277	1.355	8.836	0.277	1.404
21.966	0.679	0.956	22.030	0.677	0.958	21.983	0.681	0.723	21.874	0.684	0.959	21.883	0.686	1.054
24.392	0.754	0.538	24.535	0.754	0.580	24.436	0.757	0.559	24.305	0.760	0.601	24.308	0.762	0.585
27.530	0.851	0.192	27.692	0.851	0.189	27.567	0.854	0.194	27.407	0.857	0.207	27.402	0.859	0.210
28.403	0.878	268.8	28.603	0.879	264.8	28.471	0.882	276.3	28.302	0.885	291.4	28.295	0.887	290.8
28.759	0.889	15.40	28.928	0.889	15.51	28.794	0.892	16.39	28.622	0.895	19.80	28.614	0.897	17.58
29.600	0.915	848.6	29.807	0.916	846.7	29.633	0.918	873.4	29.454	0.921	925.2	29.412	0.922	917.1
30.668	0.948	0.834	30.946	0.951	0.892	27.503	0.852	0.831	30.509	0.954	0.902	30.465	0.955	0.751
32.350	*1.000*	*1.334*	*32.540*	*1.000*	*1.356*	*32.280*	*1.000*	*1.319*	*31.980*	*1.000*	*1.346*	*31.900*	*1.000*	*1.315*
40.470	1.251	1.626	40.675	1.250	2.830	40.641	1.259	3.770	40.711	1.273	2.919	40.704	1.276	3.536
53.313	1.648	7.710	53.593	1.647	7.869	53.649	1.662	7.754	53.662	1.678	8.138	53.688	1.683	7.539
60.850	1.881	6.084	61.013	1.875	6.029	61.009	1.890	6.031	60.986	1.907	6.812	60.993	1.912	6.609
63.115	1.951	0.792	63.225	1.943	0.936	63.204	1.958	0.996	63.224	1.977	0.929	61.503	1.928	1.024
65.832	2.035	0.072	65.959	2.027	0.091	65.948	2.043	0.102	-	-	-	-	-	-
66.156	2.045	0.185	66.284	2.037	0.266	66.303	2.054	0.301	66.295	2.073	0.265	66.958	2.099	0.304
66.803	2.065	1.442	66.967	2.058	1.831	66.949	2.074	2.074	66.966	2.094	1.769	67.596	2.119	2.084
67.644	2.091	291.2	67.878	2.086	348.5	67.788	2.100	393.1	67.798	2.120	352.0	67.788	2.125	397.0
68.032	2.103	11.76	68.204	2.096	12.13	68.175	2.112	11.85	68.213	2.133	16.85	68.202	2.138	12.73
68.323	2.112	1.296	68.464	2.104	0.850	68.466	2.121	0.956	68.501	2.142	1.083	68.713	2.154	1.154
68.711	2.124	133.6	68.952	2.119	145.3	68.853	2.133	152.8	68.853	2.153	150.9	68.872	2.159	147.5
71.591	2.213	27.04	71.718	2.204	25.37	71.758	2.223	27.27	71.763	2.244	40.42	71.265	2.234	30.33
71.752	2.218	1.326	71.913	2.21	1.635	71.920	2.228	0.778	71.923	2.249	1.769	71.903	2.254	1.937

ECD replicate analyses of HFC 134a produced by Hoechst.

Replicate 1			Replicate 2			Replicate 3			Replicate 4			Replicate 5		
Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height
2.615	0.081	62.27	2.638	0.081	85.88	2.619	0.081	92.86	2.604	0.080	94.59	2.615	0.082	90.18
3.357	0.104	1.912	3.420	0.105	1.848	3.395	0.105	2.288	3.385	0.104	1.968	3.380	0.106	2.272
6.391	0.198	82.81	6.449	0.198	82.14	6.434	0.199	82.27	6.445	0.198	81.74	6.442	0.202	85.76
6.488	0.201	2.804	-	-	-	6.531	0.202	3.456	6.543	0.201	3.904	6.537	0.205	3.936
8.619	0.267	1.090	8.664	0.266	0.768	8.664	0.268	0.800	8.691	0.267	1.296	8.674	0.272	1.600
8.845	0.274	31329	8.794	0.270	31134	8.761	0.271	32406	8.854	0.272	30814	8.706	0.273	33355
14.687	0.455	3.523	14.722	0.452	3.659	14.710	0.455	3.888	14.745	0.453	3.744	14.669	0.460	3.808
18.238	0.565	0.056	18.336	0.563	0.070	19.396	0.600	0.078	18.423	0.566	0.086	18.209	0.571	0.091
21.886	0.678	516.9	21.952	0.674	512.1	21.887	0.677	522.7	22.069	0.678	506.1	21.813	0.684	532.9
24.371	0.755	4.617	24.428	0.750	4.632	24.344	0.753	4.688	24.478	0.752	4.064	24.268	0.761	5.152
28.310	0.877	868.0	28.466	0.874	862.6	28.321	0.876	893.7	28.514	0.876	863.4	28.127	0.882	924.5
28.729	0.890	37.77	28.890	0.887	31.27	28.741	0.889	36.11	28.937	0.889	28.75	28.573	0.896	41.12
29.472	0.913	4811	29.671	0.911	4781	29.485	0.912	5010	29.751	0.914	4652	29.275	0.918	5118
-	-	-	32.472	0.997	887.4	32.201	0.996	159.8	32.420	0.996	6979	31.762	0.996	239.5
32.280	*1.000*	*1.422*	*32.570*	*1.000*	*1.369*	*32.330*	*1.000*	*1.452*	*32.550*	*1.000*	*1.798*	*31.890*	*1.000*	*1.569*
38.921	1.206	952718	39.687	1.219	946749	39.564	1.224	958766	39.386	1.210	1014071	38.972	1.222	1014071
52.487	1.626	65.95	52.633	1.616	56.06	52.666	1.629	60.00	52.633	1.617	55.94	52.650	1.651	62.05
59.879	1.855	153.4	59.896	1.839	151.5	59.907	1.853	165.7	59.925	1.841	149.0	59.921	1.879	172.2
60.299	1.868	24.01	60.352	1.853	23.03	60.360	1.867	27.57	60.348	1.854	23.62	60.368	1.893	30.496
-	-	-	-	-	-	62.947	1.947	5.104	65.621	2.016	4.128	65.662	2.059	5.248
65.625	2.033	4.585	65.563	2.013	38.46	-	-	-	-	-	-	-	-	-
65.916	2.042	2.121	-	-	-	65.630	2.030	2.464	65.914	2.025	1.600	65.949	2.068	1.952
66.755	2.068	396.2	66.769	2.050	342.3	66.826	2.067	364.5	66.760	2.051	308.2	66.778	2.094	296.74
67.175	2.081	37.70	67.159	2.062	24.84	67.182	2.078	41.62	67.183	2.064	46.03	67.160	2.106	41.34
67.820	2.101	178.2	67.843	2.083	166.5	67.828	2.098	189.8	67.834	2.084	158.6	67.862	2.128	171.6
70.725	2.191	123.5	70.709	2.171	87.03	70.738	2.188	130.1	70.731	2.173	86.66	70.764	2.219	129.2

ECD replicate analyses of HFC 134a produced by Showa Denka.

Replicate 1			Replicate 2			Replicate 3			Replicate 4			Replicate 5		
Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height
2.614	0.083	98.01	2.605	0.083	96.70	2.642	0.085	95.33	2.640	0.086	92.08	2.615	0.085	82.47
6.267	0.199	0.030	6.278	0.200	0.028	5.967	0.192	0.029	6.270	0.204	0.031	-	-	-
6.361	0.202	0.253	6.404	0.204	0.228	6.402	0.206	0.190	6.386	0.208	0.185	6.431	0.209	0.101
8.408	0.267	0.041	8.413	0.268	0.042	8.423	0.271	0.043	8.412	0.274	0.048	8.462	0.275	0.040
8.534	0.271	19.55	8.569	0.273	18.80	8.578	0.276	17.84	8.565	0.279	18.13	8.616	0.280	16.91
21.161	0.672	1.075	21.063	0.671	0.990	21.010	0.676	0.928	20.907	0.681	0.921	21.016	0.683	0.920
27.774	0.882	0.948	27.655	0.881	0.735	27.319	0.879	0.682	26.985	0.879	0.661	27.078	0.880	0.579
29.034	0.922	15.02	28.879	0.920	15.19	28.563	0.919	13.63	28.121	0.916	15.07	28.216	0.917	14.98
31.490	*1.000*	*1.421*	*31.390*	*1.000*	*1.427*	*31.080*	*1.000*	*1.471*	*30.700*	*1.000*	*1.521*	*30.770*	*1.000*	*1.540*
39.174	1.244	15.85	39.175	1.248	14.22	39.285	1.264	11.22	39.296	1.280	10.35	39.386	1.280	5.728
50.982	1.619	0.516	51.166	1.630	0.449	51.251	1.649	0.420	51.361	1.673	0.461	-	-	-
51.549	1.637	1.788	51.762	1.649	1.523	51.873	1.669	1.460	51.975	1.693	1.498	52.032	1.691	1.501
59.579	1.892	40.15	59.641	1.900	40.58	59.736	1.922	44.88	59.773	1.947	49.55	59.848	1.945	47.59
61.563	1.955	1.335	61.619	1.963	1.325	61.694	1.985	1.188	61.738	2.011	1.230	61.078	1.985	0.854
64.460	2.047	0.871	64.538	2.056	0.108	64.615	2.079	0.107	64.654	2.106	0.095	63.971	2.079	0.028
64.743	2.056	0.363	64.852	2.066	0.323	64.957	2.090	0.300	64.992	2.117	0.306	64.309	2.090	0.102
64.869	2.060	0.198	64.946	2.069	0.164	65.050	2.093	0.177	65.084	2.120	0.097	64.402	2.093	0.028
-	-	-	-	-	-	65.299	2.101	0.050	65.330	2.128	0.058	64.648	2.101	0.060
65.405	2.077	4.293	65.511	2.087	4.143	65.672	2.113	4.046	65.637	2.138	4.352	65.017	2.113	2.071
66.255	2.104	470.3	66.421	2.116	422.7	66.480	2.139	380.9	66.527	2.167	385.7	65.817	2.139	196.2
66.601	2.115	21.69	66.767	2.127	19.19	66.853	2.151	24.06	66.895	2.179	29.33	66.186	2.151	27.09
66.979	2.127	2.455	67.080	2.137	2.302	67.164	2.161	2.277	67.509	2.199	2.935	66.494	2.161	1.954
67.294	2.137	203.6	67.426	2.148	198.6	67.506	2.172	202.6	67.540	2.200	227.3	66.832	2.172	158.3
70.065	2.225	60.09	70.282	2.239	57.92	70.365	2.264	73.66	70.395	2.293	88.79	69.663	2.264	72.94
70.349	2.234	1.228	70.502	2.246	1.513	70.521	2.269	1.356	70.579	2.299	1.321	69.817	2.269	1.292
72.301	2.296	0.905	72.417	2.307	0.970	72.447	2.331	0.986	72.513	2.362	1.237	71.725	2.331	1.085

ECD replicate analyses of HFC 134a produced by Ashai Glass.

Replicate 1			Replicate 2			Replicate 3			Replicate 4			Replicate 5		
Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height
2.610	0.076	92.03	2.610	0.076	154.0	2.620	0.077	151.7	2.633	0.077	145.2	2.614	0.077	150.1
6.480	0.189	0.077	6.500	0.190	0.212	6.520	0.191	0.166	6.532	0.191	0.153	6.494	0.191	0.183
8.760	0.255	0.853	8.820	0.258	1.429	8.820	0.258	1.323	8.790	0.257	1.215	8.809	0.259	1.259
22.240	0.647	0.119	22.230	0.650	0.130	22.240	0.651	0.139	22.196	0.649	0.131	22.229	0.653	0.132
29.860	0.869	0.223	29.628	0.867	0.328	29.650	0.867	0.326	29.686	0.868	0.296	29.521	0.868	0.313
31.370	0.913	2.525	31.100	0.910	3.070	31.110	0.910	3.181	31.260	0.914	2.958	31.200	0.917	3.035
34.370	*1.000*	*1.452*	*34.180*	*1.000*	*1.396*	*34.180*	*1.000*	*1.542*	*34.201*	*1.000*	*1.649*	*34.020*	*1.000*	*1.659*
38.150	1.110	2.868	38.855	1.137	7.151	38.872	1.137	5.686	38.647	1.130	5.293	38.567	1.134	6.269
-	-	-	51.000	1.492	0.402	50.870	1.488	0.376	51.267	1.499	0.262	51.094	1.502	0.335
51.337	1.494	1.089	51.397	1.504	0.992	51.448	1.505	1.078	51.507	1.506	1.065	51.503	1.514	1.034
59.040	1.718	0.794	59.090	1.729	0.942	59.120	1.730	1.030	59.236	1.732	0.932	58.884	1.731	0.950
60.830	1.770	0.171			1.155	61.040	1.786	1.133	61.117	1.787	0.829	61.109	1.796	1.061
-	-	-	63.820	1.867	0.090	63.850	1.868	0.073	63.922	1.869	0.055	63.568	1.869	0.077
64.120	1.866	0.095	64.160	1.877	0.226	64.170	1.877	0.223	64.229	1.878	0.183	64.232	1.888	0.235
64.210	1.868	0.038	64.250	1.880	0.089	64.260	1.880	0.076	64.058	1.873	0.068	63.840	1.877	0.083
64.797	1.885	1.882				64.838	1.897	2.903	64.948	1.899	1.613	64.925	1.908	0.823
65.630	1.910	188.1	65.663	1.921	360.3	65.685	1.922	318.6	65.734	1.922	292.2	65.383	1.922	344.7
66.000	1.920	4.258	66.388	1.942	3.920	66.060	1.933	4.129	66.145	1.934	4.148	65.773	1.933	4.106
66.310	1.929	0.369				66.420	1.943	0.179	66.521	1.945	0.185	66.504	1.955	0.088
66.600	1.938	157.7	66.712	1.952	200.2	66.733	1.952	194.8	66.863	1.955	186.3	66.831	1.964	188.5
69.540	2.023	6.235	69.583	2.036	8.820	69.605	2.036	8.140	69.531	2.033	7.818	69.603	2.046	8.331

ECD replicate analyses of Air.

Replicate 1			Replicate 2			Replicate 3			Replicate 4		
Rt	*RRt*	Height	Rt	*RRt*	Height	Rt	*RRt*	Height	Rt	*RRt*	Height
2.631	0.081	248.3	2.620	0.081	246.7	2.610	0.080	251.6	2.633	0.081	237.3
6.540	0.201	0.093	6.540	0.201	0.118	6.520	0.201	0.141	6.567	0.202	0.128
8.890	0.274	0.608	8.880	0.273	0.657	8.910	0.274	0.763	8.941	0.275	0.679
32.500	*1.000*		*32.500*	*1.000*		*32.500*	*1.000*		*32.511*	*1.000*	
			34.160	1.051	0.101	34.200	1.052	0.134	34.234	1.053	0.118
35.190	1.083	0.453	35.160	1.082	0.397	35.150	1.082	0.552	35.177	1.082	0.463
39.245	1.208	34.77	39.202	1.206	30.86	39.237	1.207	37.14	39.273	1.208	32.53
50.540	1.555	0.384	50.450	1.552	0.370	50.990	1.569	0.443	50.490	1.553	0.402
51.000	1.569	0.053	50.990	1.569	0.060	51.140	1.574	0.070	51.042	1.570	0.063
			58.910	1.813	0.146	58.960	1.814	1.161	58.975	1.814	0.688
60.807	1.871	0.748	60.885	1.873	0.767	60.920	1.874	0.731	60.861	1.872	0.757
			63.720	1.961	0.066	63.750	1.962	0.092	63.917	1.966	0.079
64.677	1.990	2.957	64.900	1.997	2.990	64.708	1.991	3.144	64.762	1.992	2.935
65.510	2.016	226.7	65.910	2.028	284.4	65.940	2.029	306.5	66.030	2.031	282.7
			66.220	2.038	0.161	66.290	2.040	0.129	66.322	2.040	0.141
66.530	2.047	176.9	66.545	2.048	183.5	66.592	2.049	180.5	66.680	2.051	181.1
69.400	2.135	5.662	69.400	2.135	5.323	69.490	2.138	6.020	69.346	2.133	5.615

ECD replicate analyses of Helium.

Replicate 1			Replicate 2			Replicate 3			Replicate 4		
Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height	Rt	RRt	Height
8.507	0.262	0.529	8.580	0.264	0.564	8.592	0.264	0.548	8.635	0.266	0.330
32.500	1.000		32.500	1.000		32.500	1.000		32.500	1.000	
34.410	1.059	0.398	34.360	1.057	0.335	34.510	1.062	0.323	34.710	1.068	0.229
38.475	1.184	6.678	38.597	1.188	5.771	38.678	1.190	4.122	38.808	1.194	3.841
			60.200	1.852	0.376	60.180	1.852	0.346	60.455	1.860	0.566
65.107	2.003	9.969	65.260	2.008	9.414	65.293	2.009	8.404	65.392	2.012	8.672
66.053	2.032	56.36	66.205	2.037	58.75	66.265	2.039	67.02	66.382	2.043	74.68

Appendix I

Mass spectrometry analysis of ICI sample produced in UK

Scan number	Retention time (min:sec)	Ions present	Identity	Possible identity
182	3:06	44, 40, 34	Air	
609	10:19	69, 65, 45	HFC 143a	
696	11:47	114, 95, 75, 64, 45		Isomer of HFC 1234
721	12:12	135*, 101, 85*, 50	CFC 114	
727	12:18	151*, 135*, 101*, 85*, 69, 66	CFC 114a	
740	12:32	96, 95, 77, 51, 46, 44	HFC 1243zf	
747	12:39	65, 51, 45, 44		HFC 152
750	12:48	101, 100, 83, 69, 51, 33	HFC 134a	
880	14:54	118*, 101, 83, 69, 63, 51, 33		Chlorinated HFC 134a
1000	16:56	136*, 117*, 101, 67*, 51	CFC 124	
1013	17:09	101, 83, 63, 51, 33	HFC 134	

* indicates presence of chlorine isotope ratio in mass spectrum.

Mass spectrometry analysis of ICI sample produced in USA.

Scan number	Retention time (min:sec)	Ions present	Identity	Possible identity
182	3:07	44, 40, 34	Air	
211	3:34	119, 100, 69, 50	HFC 116	
291	4:56	64, 50, 45, 44, 33	1122a	
583	9:52	131, 119, 69		HFC 218 or HFC 227
584	9.53	169, 150, 131, 119, 100, 69, 50		HFC 218 or HFC 227
587	9:57	169, 150, 131, 119, 100, 69, 50		HFC 218 or HFC 227
610	10:20	83, 82, 69, 66, 51, 45, 33	HFC 143a	HFC 152
647	10:57	185, 133, 119, 100, 69, 50, 47		
715	12:06	101, 69, 51	HFC 125	
759	12:65	101, 100, 83, 69, 51, 33	HFC 134a	
885	14:59	118*, 99, 83, 69, 51, 33		Chlorinated HFC 134a
1003	16:59	151, 82, 69, 51		HFC 226

* indicates presence of chlorine isotope ratio in mass spectrum.

Mass spectrometry analysis of ICI sample produced in Japan

Scan number	Retention time (min:sec)	Ions present	Identity	Possible identity
182	3:07	44, 40, 34	Air	
678	11:28	98*, 79*, 71, 63, 48	CFC 1122	
720	12:11	133, 106, 96, 77, 69, 51		HFC 1243
725	12:15	101, 100, 83, 69, 51, 33	HFC 134a	
883	14:56	118*, 99, 83, 69, 63, 51, 33		Chlorinated HFC 134a
989	16:44	117*, 101, 85, 67, 51	CFC 124	
1001	16:56	136*, 117*, 101, 67*, 51		CFC 124
1014	17:09	101, 100, 83, 51, 33	HFC 134	
1086	18:23	151, 118*, 99^{+}, 83, 49	CFC 123	
1093	18:30	118*, 99, 83, 69, 49, 33		CFC 133

* indicates presence of chlorine isotope ratio in mass spectrum.

$^{+}$ indicates presence of two chlorine atoms due to isotope ratio in mass spectrum.

Mass spectrometry analysis of Atochem sample of HFC 134a

Scan number	Retention time (min:sec)	Ions present	Identity	Possible identity
182	3:06	44, 40, 34	Air	
683	11:34	98*, 79, 63, 48	CFC 1122	
688	11:39	133, 114, 95, 75, 69, 64, 45	HFC 1234	
712	12:03	151*, 135*, 101, 85*, 50	CFC 114	
7217	12:08	151*, 135*, 101*, 85*, 69, 66	CFC 114a	
721	12:12	133, 98*, 83, 63, 56, 43		Isomer of CFC 1122
728	12:20	96, 95, 83, 77, 51	HFC 1243zf	
733	12:25	119, 95, 69, 65, 45, 44, 33		HFC 245cb
734	12:27	101, 83, 69, 63, 51, 33	HFC 134a	
884	14:58	118*, 101, 83, 69, 63, 51, 33		Chlorinated HFC 134a
999	16:55	136*, 117*, 101, 67*, 51	CFC 133	
1011	17:09	101, 83, 63, 51, 33	HFC 134	

* indicates presence of chlorine isotope ratio in mass spectrum.

Mass spectrometry analysis of Hoechst sample of HFC 134a

Scan number	Retention time (min:sec)	Ions present	Identity	Possible identity
182	3:07	44, 40, 34	Air	
465	7:52	120, 101*, 85*, 66, 50, 35*	CFC 12	
602	10:11	69, 65, 45	HFC 143a	
606	10:16	162, 150, 131, 112, 100, 74, 69, 50		C_4F_8
665	11:16	185*, 169, 135*, 116*, 85*, 69, 50		Isomer of CFC 217
680	11:31	98*, 79*, 63, 48*, 44	CFC 1122	
685	11:36	114, 95, 83, 69, 64, 45	HFC 1234yf	
696	11:47	119, 101, 69, 51	HFC 125	
700	11:51	85*, 67, 51, 35*	CFC 22	
707	11:58	151*, 135, 116, 101, 85*, 50, 35*	CFC 114	
713	12:04	151^+, 135*, 101^+, 85*, 69, 66, 35	CFC 114a	
717	12:08	135, 98*, 85, 69, 63, 43		
723	12:14	106, 96, 77, 51, 33	1243	
728	12:20	65, 45, 44		HFC 125
729	12:22	101, 100, 83, 69, 63, 51, 33	HFC 134a	
884	14:58	118*, 99, 83, 69, 63, 51, 33		Chlorinated HFC 134a
999	16:55	136*, 117*, 101, 67, 51	CFC 133a	
1013	17:09	101, 83, 63, 51, 33	HFC 134	
1091	18:28	118*, 99, 83, 69, 64, 49	CFC 133	

* indicates presence of chlorine isotope ratio in mass spectrum.

Mass spectrometry analysis of Showa Denka sample of HFC 134a

Scan number	Retention time (min:sec)	Ions present	Identity	Possible identity
184	3:10	44, 40, 34	Air	
607	10:16	69, 65, 45, 33	HFC 143a	
726	12:06	101, 100, 83, 69, 51, 33	HFC 134a	
890	15:04	118*, 99, 83, 69, 33		Chlorinated HFC 134a
989	16:44	136, 117, 101,85, 67		
1002	16:57	136*, 117*, 101, 67, 51	CFC 133	
1016	17:12	117, 101*, 82, 67, 51	CFC 124	
1088	18:25	118*, 99, 83, 69, 49, 33	CFC 123	
1095	18:32	118*, 99, 83, 69, 49, 33	CFC 133	

*indicates presence of chlorine isotope ratio in mass spectrum.

Mass spectrometry analysis of Ashai Glass sample of HFC 134a

Scan number	Retention time (min:sec)	Ions present	Identity	Possible identity
182	3:06	44, 40, 34	Air	
606	10:15	82, 69, 65, 45, 33	HFC 143a	
750	12:48	101, 100, 83, 69, 51, 33	HFC 134a	
990	16:45	117, 101, 85, 67		
1000	16:55	136*, 117*, 101, 67*, 51	CFC 133	
1012	17:07	101, 83, 51, 33	HFC 134	
1086	18:23	209, 208, 191, 151*, 118, 101, 83, 63, 49	CFC 123	
1092	18:29	118*, 99*, 83, 49, 33	CFC 133	

* indicates presence of chlorine isotope ratio in mass spectrum.

Mass spectrometry analysis of Unknown sample labelled 98Q of HFC 134a.

Scan number	Retention time (min:sec)	Ions present	Identity	Possible identity
182	3:06	44, 40, 34	Air	
613	10:23	83, 82, 69, 65, 45, 33	HFC 143a	
676	11:49	133, 98*, 79, 63, 56, 48*	CFC 1122	
680	11:31	114, 95, 75, 69, 64, 45	1234	
688	11:39	129^, 85*, 79^, 50	12B1	
719	12:10	101, 91, 84, 69, 67, 51, 44	CFC 125	
731	12:23	151^{+}, 135*, 116, 101*, 85*, 50, 35	CFC 114	
737	12:29	151^{+}, 135*, 101^{+}, 85*, 69, 47, 35	CFC 114a	
752	12:44	96, 95, 77, 75, 51, 46, 40	HFC 1243zf	
765	12:52	101, 100, 83	HFC 134a	
873	14:54	118*, 101, 83, 69, 63, 51, 33		Chlorinated HFC 134a
1002	16:58	136*, 117*, 101, 67*, 51, 35	CFC 133	
1013	17:09	101, 83, 63, 51, 33	HFC 134	

* indicates presence of chlorine isotope ratio in mass spectrum.

$^{+}$ indicates presence of two chlorine atoms as seen in isotopic mass ratio.

^ indicates presence of bromine isotope ratio in mass spectrum.

Mass spectrometry analysis of unknown sample labelled Q red of HFC 134a

Scan number	Retention time (min:sec)	Ions present	Identity	Possible identity
183	3:06	44, 40, 34	Air	
395	6:41	135*, 119, 85*, 50	CFC 115	
604	10:14	69, 65, 64, 45, 33	HFC 143a	
682	11:33	98*, 79*, 63, 48	1122	
699	11:50	101, 69, 51	HFC 125	
711	12:02	135*, 101*, 85*, 50	CFC 114	
715	12:06	151^{+}, 135*,101^{+}, 85*, 69, 66, 50	CFC 114a	
720	12:11	98*, 79, 63, 48, 44		
732	12:25	101, 100, 83, 69, 63, 51, 33	HFC 134a	
885	14:59	118*, 99, 83, 69, 63, 51, 33		Chlorinated HFC 134a
1001	16:57	136*, 117*, 101, 67*, 51	CFC 133	
1013	17:09	101, 83, 63, 51, 33	HFC 134	
1093	18:30	118*, 99*, 83, 69, 49*, 33	CFC 133	

* indicates presence of chlorine isotope ratio in mass spectrum.

$^{+}$ indicates presence of two chlorine atoms as seen in isotopic ratio

Mass spectrometry analysis of unknown sample labelled Q of HFC 134a

Scan number	Retention time (min:sec)	Ions present	Identity	Possible identity
182	3:06	44, 40, 34	Air	
613	10:23	52, 51, 33	HFC 32	
674	11:25	98*, 79*, 63, 50, 48	CFC 1122a	
700	11:51	135*, 101, 85*, 66, 50	CFC 114	
704	11:55	151*, 135*, 101^{+}, 85*, 69, 66	CFC 114a	
713	12:04	96, 95, 83, 77, 51, 44	CFC 1243	
717	12:11	101, 100, 83, 69, 51, 33	HFC 134a	
877	14:51	118*, 101, 83, 69, 63, 51, 33		Chlorinated HFC 134a
1002	16:58	136*, 117*, 101, 67*, 51	CFC 133	
1015	17:11	101, 83, 63, 51, 33	HFC 134	

* indicates presence of chlorine isotope ratio in mass spectrum.

$^{+}$ indicates presence of two chlorine atoms as seen in isotopic ratio.

Appendix J

Mixed samples of HFC 134a analysed by GC FID

UK/USA 50:50 mixed sample			Height response (mV)		Ratio compared to mixed sample	
Retention time (min)	RRT	Peak height (mV)	UK expected	USA expected	UK ratio	USA ratio
2.65	0.086	0.65	0.651	0.539	1.00	1.21
3.05	0.099	0.47	-	1.034	-	0.45
4.26	0.138	3.16	-	7.793	-	0.41
4.84	0.157	0.08	-	0.163	-	0.49
13.05	0.423	1.79	-	4.739	-	0.38
13.14	0.426	18.61	-	35.005	-	0.53
13.32	0.432	17.28	3.312	41.086	5.22	0.42
-	-	0.0569	-	0.124	-	0.46
17.74	0.578	0.236	-	0.486	-	0.49
-	-	-	-	0.077	-	NS
19.55	0.634	0.04		0.063	-	0.70
-	-	-	-	0.055	-	NS
-	-	0.2656	0.483	-	0.55	-
-	-	-	0.081	-	NS	-
26.43	0.857	1.07	0.064	2.356	16.81	0.45
27.42	0.889	1.12	1.765	-	0.64	-
28.56	0.926	1.93	3.240	-	0.60	-
29.54	0.958	0.64	-	0.140	-	4.56
29.70	0.963	0.48	0.795	-	0.61	-
30.69	0.995	0.32	0.527	-	0.61	-
30.84	1.000	OS	OS	OS	NA	NA
50.58	1.640	4.39	7.447	2.837	0.59	1.55
51.75	1.678	1.52	2.507	0.536	0.61	2.83
-	-	-	-	3.302	NA	-
58.97	1.912	0.27	0.396	-	0.67	-
65.78	2.133	4.73	11.684	4.044	0.40	1.17
66.86	2.168	0.16	-	1.613	-	0.10
68.22	2.212	1.74	3.616	1.929	0.48	0.90
71.46	2.317	2.45	1.745	-	1.40	-

Mixed samples of HFC 134a analysed by GC FID

UK/Japan 50:50 mixed sample			Height response (mV)		Ratio compared to mixed sample	
Retention time (min)	RRT	Peak height (mV)	UK expected	Japan expected	UK ratio	Japan ratio
2.67	0.084	0.83	0.651	0.589	1.27	1.40
13.60	0.427	1.84	3.312	-	0.55	-
23.05	0.724	0.27	0.483	-	0.55	-
24.17	0.759	0.15	0.081	0.241	1.80	0.61
27.19	0.854	0.05	0.064	-	-	-
28.18	0.885	1.05	1.765	0.206	0.59	5.06
29.29	0.920	1.80	3.240	-	0.56	-
30.53	0.959	0.51	0.795	0.271	0.64	1.87
31.62	0.993	0.31	0.527	-	0.58	-
31.84	1.000	OS	OS	OS	-	-
45.31	1.423	3.13	-	-	-	-
51.20	1.608	4.74	7.447	0.706	0.64	6.71
52.60	1.652	1.43	2.507	-	0.57	-
58.94	1.851	1.63	-	2.522	-	0.64
65.75	2.065	0.36	0.396	-	0.91	-
66.83	2.099	7.27	11.684	5.123	0.62	1.42
68.46	2.150	0.29	-	0.136	-	2.13
71.42	2.243	0.13	-	0.259	-	0.49
73.30	2.302	1.17	3.616	1.933	0.32	0.61
73.96	2.323	1.69	1.745	1.700	0.97	0.99

Mixed samples of HFC 134a analysed by GC FID

USA/Japan 50:50 mixed sample			Height response (mV)		Ratio compared to mixed sample	
Retention time (min)	RRT	Peak height (mV)	USA expected	Japan expected	USA ratio	Japan ratio
2.66	0.08	0.505	0.539	0.589	0.937	0.857
3.08	0.10	0.495	1.034	-	0.479	-
4.27	0.13	3.885	7.793	-	0.498	-
4.88	0.15	0.080	0.163	-	0.490	-
13.16	0.41	2.184	4.739	-	0.461	-
13.29	0.41	20.296	35.005	-	0.580	-
13.48	0.42	18.412	41.086	-	0.448	-
14.77	0.46	0.072	0.124	-	0.581	-
17.98	0.56	0.243	0.486	-	0.500	-
18.49	0.58	0.051	0.077	-	0.662	-
19.90	0.62	0.050	0.063	-	0.799	-
21.80	0.68	0.041	0.055	-	0.751	-
24.27	0.76	0.111	0.241	-	0.461	-
27.29	0.85	1.263	2.356	-	0.536	-
28.47	0.89	0.100	-	0.206	-	0.484
30.72	0.96	0.161	0.140	-	1.148	-
31.14	0.97	0.122	-	0.271	-	0.450
32.10	1.00	OS	OS	OS	NA	NA
47.35	1.48	0.098	-	-	-	-
48.82	1.52	0.215	-	-	-	-
50.72	1.58	0.320	-	-	-	-
53.29	1.66	2.363	2.837	0.706	0.833	3.349
54.51	1.70	0.308	0.536	-	0.575	-
65.74	2.05	3.545	3.302	2.522	1.074	1.406
66.80	2.08	0.099	-	-	-	-
68.57	2.14	0.554	4.044	5.123	0.137	0.108
-	-	-	-	0.136	-	NS
71.45	2.23	2.080	1.613	0.259	1.290	8.031
73.28	2.28	1.416	1.929	1.933	0.734	0.733

Mixed samples of HFC 134a analysed by GC FID

USA/Ashai Glass 20:80 mixed sample			Height response (mV)		Ratio compared to mixed sample	
Retention time (min)	RRT	Peak height (mV)	USA expected	Ashai expected	USA ratio	Ashai ratio
2.696	0.082	0.620	0.539	0.582	1.15	1.07
3.058	0.093	0.146	1.034	-	0.14	-
4.242	0.129	0.801	7.793	-	0.10	-
-	-	-	0.163	-	-	-
13.053	0.397	0.567	4.739	0.656	0.12	0.86
13.251	0.403	7.006	35.005	-	0.20	-
13.415	0.408	7.964	41.086	-	0.19	-
-	-	-	0.124	-	-	-
17.821	0.542	0.062	0.486	-	0.13	-
-	-	-	0.077	-	-	-
-	-	-	0.063	-	-	-
-	-	-	0.055	-	-	-
26.994	0.821	0.319	2.356	-	0.14	-
-	-	-	0.140	-	-	-
32.880	1.000	OS	OS	OS	NA	NA
46.361	1.410	1.982	2.837	1.419	0.70	1.40
50.109	1.524	1.221	0.536	0.170	2.28	7.19
59.381	1.806	0.146	3.302	-	0.04	-
65.102	1.980	5.980	4.044	2.733	1.48	2.19
65.760	2.000	.0.986	-	0.115	-	-
66.681	2.028	0.283	1.613	-	0.18	-
71.251	2.167	4.565	1.929	1.983	2.37	2.30
72.435	2.203	2.973	-	2.672	-	1.11

Mixed samples of HFC 134a analysed by GC FID

UK/Atochem 25:75 mixed sample			Height response (mV)		Ratio compared to mixed sample	
Retention time (min)	RRT	Peak height (mV)	UK expected	Atochem expected	UK ratio	Atochem ratio
2.66	0.08	0.62	0.651	0.566	0.96	1.10
13.47	0.42	0.82	3.312	0.095	0.25	8.62
16.42	0.51	0.04	-	-	-	-
23.05	0.71	0.58	0.483	0.881	1.20	0.66
24.18	0.75	0.39	0.081	0.736	4.74	0.52
-	-	-	0.064	-	NS	-
28.34	0.87	0.95	1.765	1.093	0.54	0.87
29.57	0.91	1.09	3.240	0.743	0.34	1.46
30.58	0.94	0.09	-	0.113	-	0.75
30.80	0.95	0.30	0.795	0.321	0.38	0.94
32.14	0.99	1.00	0.527	1.832	1.89	0.54
32.46	1.00	OS	OS	OS	NA	NA
45.44	1.40	0.06	-	-	-	-
49.86	1.54	14.17	7.447	15.525	1.90	0.91
51.32	1.58	1.29	2.507	0.938	0.51	1.37
52.84	1.63	1.04	-	-	-	-
58.75	1.81	0.37	0.396	0.485	0.94	0.77
61.02	1.88	0.08	-	-	-	-
61.28	1.89	0.14	-	-	-	-
64.79	2.00	0.15	-	-	-	-
65.67	2.02	9.71	11.684	2.015	0.83	4.82
66.67	2.05	0.29	-	-	-	-
71.22	2.19	5.34	3.616	1.773	1.48	3.01
76.38	2.353	0.365	1.745	-	0.21	-

Mixed samples of HFC 134a analysed by GC FID

UK/Hoechst 60:40 mixed sample			Height response (mV)		Ratio compared to mixed sample	
Retention time (min)	RRT	Peak height (mV)	UK expected	Hoechst expected	UK ratio	Hoechst ratio
2.647	0.089	0.579	0.651	0.489	0.889	1.183
3.033	0.102	0.050	-	-	-	-
3.361	0.113	0.050	-	0.132	-	0.380
3.420	0.115	0.644	-	2.952	-	0.218
3.539	0.119	0.050	-	0.119	-	0.421
-	-	-	-	0.078	-	-
6.364	0.214	0.073	-	0.149	-	0.489
7.227	0.243	0.080	-	0.197	-	0.406
7.762	0.261	0.100	-	0.515	-	0.194
8.535	0.287	2.181	-	8.303	-	0.263
12.907	0.434	9.483	3.312	24.310	2.864	0.390
14.245	0.479	6.784	-	23.279	-	0.291
16.654	0.560	0.070	-	-	-	-
20.907	0.703	0.396	-	1.730	-	0.229
22.127	0.744	0.464	0.483	0.343	0.961	1.353
23.227	0.781	1.179	0.081	4.775	14.520	0.247
25.993	0.874	0.224	0.064	0.760	3.522	0.295
26.736	0.899	5.105	1.765	7.028	2.893	0.726
27.688	0.931	4.039	-	7.900	-	0.511
-	-	-	3.240	4.199	-	-
28.699	0.965	0.093	-	0.206	-	0.452
28.848	0.970	0.723	0.795	0.409	0.909	1.768
29.562	0.994	0.053	-	0.118	-	0.448
29.710	0.999	0.593	0.527	0.661	1.126	0.898
29.740	1.000		OS	OS	-	-
49.398	1.661	6.156	7.447	8.395	0.827	0.733
50.796	1.708	3.721	2.507	2.285	1.484	1.628
58.350	1.962	5.359	-	19.650	-	0.273
-	-	-	0.396	5.145	-	-
65.398	2.199	3.189	11.684	2.386	0.273	1.337
70.960	2.386	2.951	3.616	1.724	0.816	1.711
71.168	2.393	1.039	1.745	-	0.595	-

Mixed samples of HFC 134a analysed by GC FID

USA/Hoechst 50:50 mixed sample			Height response (mV)		Ratio compared to mixed sample	
Retention time (min)	RRT	Peak height (mV)	USA expected	Hoechst expected	USA ratio	Hoechst ratio
2.656	0.090	0.328	0.539	0.489	0.61	0.67
3.040	0.103	0.405	1.034	0.132	0.39	3.08
3.335	0.113	0.107	-	-	-	-
3.394	0.115	1.639	-	2.952	-	0.56
3.541	0.120	0.053	-	0.119	-	0.45
4.190	0.142	3.470	7.793	0.078	0.45	44.26
4.840	0.164	0.063	0.163	-	0.39	-
6.404	0.217	0.084	-	0.149	-	0.56
7.318	0.248	0.122	-	0.197	-	0.62
7.820	0.265	0.289	-	0.515	-	0.56
8.676	0.294	5.182	-	8.303	-	0.62
12.896	0.437	1.995	4.739	-	0.42	-
13.014	0.441	20.002	35.005	-	0.57	-
13.073	0.443	40.302	41.086	24.310	0.98	1.66
14.342	0.486	15.085	0.124	23.279	121.65	0.65
17.440	0.591	0.192	0.486	-	0.39	-
17.913	0.607	0.053	0.077	-	0.69	-
-	-	-	0.063	-	NS	-
20.893	0.708	1.026	0.055	1.730	18.79	0.59
22.073	0.748	0.180	-	0.343	-	0.52
23.077	0.782	2.975	-	4.775	-	0.62
25.733	0.872	1.729	2.356	0.760	0.73	2.28
26.441	0.896	5.407	-	7.028	-	0.77
26.589	0.901	5.932	-	7.900	-	0.75
27.562	0.934	2.969	-	4.199	-	0.71
28.507	0.966	0.119	0.140	0.206	0.85	0.58
28.684	0.972	0.334	-	0.409	-	0.82
29.185	0.989	0.041	-	0.118	-	0.35
29.333	0.994	0.091	-	-	-	-
29.451	0.998	0.381	-	0.661	-	0.58
29.510	1.000	OS	OS	OS	NA	NA
51.318	1.739	5.396	2.837	8.395	1.90	0.64
53.472	1.812	1.234	0.536	2.285	2.30	0.54
62.886	2.131	9.9485	3.302	19.650	3.01	0.51
62.827	2.129	2.321		5.145	-	0.45
64.774	2.195	3.564	4.044	2.386	0.88	1.49
69.909	2.369	1.123	1.613		0.70	-
72.418	2.454	1.665	1.929	1.724	0.86	0.97

Appendix K

Mixed samples of HFC 134a analysed by GC ECD

USA/Japan 50:50 mixed sample			Height response (mV)		Ratio compared to mixed sample	
Retention time (min)	RRT	Peak height (mV)	USA expected	Japan expected	USA ratio	Japan ratio
2.64	0.08	127.2	120.2	114.1	1.1	1.1
3.11	0.09	46.98	156.8	0.03	0.3	1446
3.38	0.10	0.12	0.28	-	0.4	-
5.42	0.17	0.03	0.10	-	0.3	-
6.21	0.19	0.41	1.07	-	0.4	-
6.32	0.19	0.54	0.29	0.66	1.9	0.8
8.45	0.26	30.50	16.51	40.00	1.8	0.8
12.65	0.39	5.36	15.90	-	0.3	-
12.78	0.39	0.20	-	-	-	-
15.55	0.47	0.06	0.20	-	0.3	-
17.28	0.53	0.06	0.17	-	0.3	-
20.90	0.64	6.49	13.40	1.63	0.5	4.0
23.08	0.70	0.05	-	0.12	-	0.4
26.21	0.80	0.11	0.22	0.04	0.5	3.0
26.69	0.81	8.21	22.20	-	0.4	-
27.44	0.84	0.13	-	-	-	-
28.05	0.86	0.09	1.60	-	0.1	-
28.32	0.86	0.38	1.89	0.35	0.2	1.1
-	-	-	-	0.07	-	NS
29.53	0.90	6.93	10.89	5.20	0.6	1.3
30.63	0.93	0.05	-	0.09	-	0.6
32.77	1.00	1.50	1.55	1.40	1.0	1.1
38.11	1.16	10.00	21.50	3.96	0.5	2.5
47.31	1.44	49.54	154.4	-	0.3	-
48.69	1.49	0.11	-	-	-	-
49.65	1.52	0.57	-	-	-	-
58.24	1.78	0.11	-	-	-	-
60.30	1.84	1.52	1.08	0.90	1.4	1.7
63.82	1.95	0.32	-	0.07	-	4.7
63.75	1.95	0.18	0.26	0.24	0.7	0.8
64.32	1.96	5.92	3.17	1.99	1.9	3.0
65.21	1.99	416.4	400.7	340.1	1.0	1.2
65.53	2.00	2.77	4.39	3.30	0.6	0.8
65.88	2.01	0.84	-	0.25	-	3.3
66.19	2.02	268.1	181.0	132.2	1.5	2.0
69.05	2.11	10.26	10.08	7.86	1.0	0.8

Mixed samples of HFC 134a analysed by GC ECD.

UK/Atochem 50:50 mixed sample			Height response (mV)		Ratio compared to mixed sample	
Retention time (min)	RRT	Peak height (mV)	UK expected	Atochem expected	UK ratio	Atochem ratio
2.63	0.083	80.95	115.6	72.12	0.70	1.12
3.07	0.097	0.05	-	-	-	-
6.24	0.198	0.06	0.14	-	0.42	-
6.34	0.201	0.56	0.97	0.21	0.58	2.65
8.33	0.264	0.07	0.18	-	0.40	-
8.48	0.269	1.47	3.12	1.31	0.47	1.12
20.77	0.659	35.85	79.10	-	0.45	-
22.03	0.699	0.24	-	0.93	-	0.26
23.51	0.746	0.23	-	0.57	-	0.40
26.24	0.832	0.28	0.47	0.20	0.60	1.41
27.06	0.859	350.1	493.3	278.4	0.71	1.26
27.38	0.869	512.0	1058	16.94	0.48	30.2
28.39	0.901	1903	3605	882.2	0.53	2.16
29.50	0.936	1.35	1.92	0.84	0.70	1.60
31.52	*1.000*	*1.37*	*1.55*	*1.33*	*0.88*	*1.02*
38.23	1.213	2.00	10.33	2.94	0.19	0.68
49.01	1.555	0.58	1.40	-	0.41	-
50.40	1.599	13.95	22.10	7.80	0.63	1.79
58.03	1.841	3.83	9.33	6.31	0.41	0.61
60.44	1.918	1.30	1.48	0.94	0.88	1.39
63.50	2.015	0.56	-	0.09	-	6.34
63.81	2.024	0.14	-	0.26	-	0.53
63.92	2.028	1.36	3.68	-	0.37	-
64.51	2.047	5.84	5.47	1.84	1.07	3.17
65.38	2.074	360.0	445.1	356.4	0.81	1.01
65.72	2.085	7.75	3.41	13.06	2.27	0.59
66.03	2.095	0.53	-	1.07	-	0.50
66.39	2.106	261.1	189.6	146.0	1.38	1.79
69.21	2.196	21.84	12.08	30.09	1.81	0.73
70.6	2.24	0.69	-	1.49	-	0.46

Mixed samples of HFC 134a analysed by GC ECD.

USA/Showa 50/50 mixed sample			Height response (mV)		Ratio compared to mixed sample	
Retention time (min)	RRT	Peak height (mV)	USA expected	Showa expected	USA ratio	Showa ratio
2.62	0.08	77.60	120.2	92.92	0.65	0.84
3.09	0.10	69.78	156.8	-	0.44	-
3.36	0.11	0.14	0.28	-	0.50	-
5.70	0.18	0.05	0.10	-	0.55	-
6.21	0.20	0.53	1.07	0.03	0.49	17.80
6.30	0.20	0.19	0.29	0.19	0.65	0.98
8.27	0.27	0.04	-	0.04	-	0.93
8.43	0.27	15.57	16.51	18.25	0.94	0.85
12.57	0.41	7.35	15.90	-	0.46	-
15.42	0.50	0.10	0.20	-	0.48	-
17.10	0.55	0.10	0.17	-	0.58	-
20.55	0.66	7.74	13.40	0.97	0.58	8.01
25.45	0.82	0.16	0.22	-	0.73	-
25.94	0.84	10.99	22.20	-	0.49	-
27.00	0.87	0.24	1.60	-	0.15	-
27.27	0.88	0.49	1.89	0.72	0.26	0.68
28.28	0.91	6.87	10.89	14.78	0.63	0.46
31.01	1.00	1.45	1.55	1.48	0.94	0.98
38.11	1.23	10.00	21.50	11.47	0.47	0.87
47.51	1.53	59.52	154.4	-	0.39	-
49.80	1.61	0.28	-	0.46	-	0.61
50.44	1.63	0.78	-	1.55	-	0.50
58.35	1.88	17.44	1.08	44.55	16.12	0.39
60.31	1.94	1.24	-	1.19	-	1.05
61.85	1.99	0.09	0.26	0.24	0.36	0.38
62.24	2.01	0.15	-	0.28	-	0.54
63.40	2.04	0.10	-	0.13	-	0.75
63.71	2.05	0.18	-	0.06	-	3.21
64.41	2.08	4.51	3.17	3.78	1.42	1.19
65.30	2.11	295.2	400.7	371.2	0.74	0.80
65.68	2.12	12.49	4.39	24.27	2.85	0.51
65.99	2.13	1.20	-	2.38	-	0.50
66.30	2.14	256.2	181.0	198.1	1.42	1.29
69.13	2.23	37.16	10.08	70.68	3.69	0.53
69.29	2.23	0.53	-	1.34	-	0.39
71.20	2.30	0.48	-	1.04	-	0.46

Mixed samples of HFC 134a analysed by GC ECD.

UK/Hoechst 80/20 mixed sample			Height response (mV)		Ratio compared to mixed sample	
Retention time (min)	RRT	Peak height (mV)	UK expected	Hoechst expected	UK ratio	Hoechst ratio
2.61	0.08	113.7	115.6	85.16	0.98	1.34
3.35	0.11	0.4	-	2.06	-	0.19
6.23	0.20	17.66	0.14	82.94	124	0.21
6.31	0.20	1.02	0.97	3.53	1.05	0.29
8.31	0.26	0.31	0.18	1.11	1.77	0.28
8.53	0.27	6025	3.12	31808	1928	0.19
14.18	0.45	0.83	-	3.72	-	0.22
-	-	-	-	0.08	-	NS
20.77	0.66	160.9	79.10	518.1	2.03	0.31
21.94	0.69	0.06	-	-	-	-
22.93	0.73	0.95	-	4.63	-	0.21
26.24	0.83	0.28	0.47	-	0.60	-
26.6	0.84	0.64	-	-	-	-
26.65	0.84	0.65	-	-	-	-
27.06	0.86	574.7	493.3	882.4	1.17	0.65
27.4	0.87	782.2	1058	35.00	0.74	22.4
28.38	0.90	2697	3605	4874	0.75	0.55
31.28	0.99	336.9	1.92	2066	175	0.16
31.60	1.00	1.465	1.55	1.52	0.95	0.96
38.00	1.20	OS	10.33	977275	NR	NR
49.15	1.56	0.97	1.40	-	0.69	-
50.52	1.60	29.7	22.10	60.00	1.34	0.50
58.1	1.84	47.21	-	158.4	-	0.30
58.55	1.85	4.09	9.33	25.75	0.44	0.16
60.54	1.92	2.19	1.48	4.83	1.48	0.45
63.91	2.02	0.11	-	-	-	-
63.97	2.02	4.36	-	21.52	-	0.20
64.26	2.03	0.91	3.68	-	0.25	-
64.58	2.04	2.11	5.47	2.03	0.39	1.04
65.46	2.07	143.8	445.1	341.6	0.32	0.42
65.81	2.08	10.59	-	38.31	-	0.28
66.44	2.10	0.62	3.41	-	0.18	-
66.46	2.10	187.4	189.6	172.9	0.99	1.08
69.28	2.19	25.06	12.08	111.3	2.07	0.23

Mixed samples of HFC 134a analysed by GC ECD.

USA/Ashai 50/50 mixed sample			Height response (mV)		Ratio compared to mixed sample	
Retention time (min)	RRT	Peak height (mV)	USA expected	Ashai expected	USA ratio	Ashai ratio
2.62	0.08	126.6	120.2	138.6	1.05	0.91
3.08	0.10	59.96	156.8	-	0.38	-
3.34	0.10	0.18	0.28	-	0.65	-
-	-	-	0.10	-	N/S	-
6.14	0.19	0.6	1.07	-	0.56	-
6.23	0.19	0.08	0.29	0.16	0.28	0.51
8.32	0.26	8.02	16.51	1.22	0.49	6.60
12.44	0.39	7.17	15.90	-	0.45	-
15.29	0.48	0.06	0.20	-	0.30	-
16.96	0.53	0.07	0.17	-	0.37	-
20.44	0.64	10.04	13.40	-	0.75	-
21.01	0.65	0.06	-	0.13	-	0.49
25.10	0.78	0.11	0.22	-	0.49	-
25.86	0.80	10.04	22.20	-	0.45	-
26.99	0.84	0.11	-	-	-	
27.24	0.85	0.48	1.60	-	0.30	-
28.32	0.88	2.05	1.89	-	1.08	-
28.95	0.90	5.33	10.89	0.30	0.49	17.9
29.44	0.91	1.36	-	2.95	-	0.46
32.17	1.00	*1.46*	1.55	1.54	0.94	0.95
37.96	1.18	5.08	21.50	5.45	0.24	0.93
47.13	1.47	62.87	154.4	0.34	0.41	183
49.78	1.55	0.37	-	1.05	-	0.35
57.92	1.80	0.33	1.08	0.93	0.31	0.35
59.64	1.85	0.32	-	-	-	-
60.11	1.87	0.78	0.26	0.87	3.04	0.90
63.63	1.98	0.12	-	0.07	-	1.68
63.74	1.98	0.07	-	0.19	-	0.36
63.91	1.99	0.06	-	0.07	-	0.78
64.35	2.00	3.44	3.17	1.81	1.09	1.91
65.24	2.03	169.5	400.7	300.8	0.42	0.56
65.52	2.04	1.93	4.39	4.11	0.44	0.47
-	-	-	-	0.21	-	NS
66.22	2.06	222.7	181.0	185.5	1.23	1.20
69.00	2.14	6.18	10.08	7.87	0.61	0.79

Appendix L

Comparison of peaks (denoted by scan number) between samples of HFC 134a based on mass spectral data (both for known and unknown spectra).

UK	USA	Japan	Atochem	Hoechst	Showa Denka	Ashai Glass	Peak identity
182	182	182	182	182	184	182	Air
	212						HFC 116
	291						*HFC 1132*
				465			CFC 12
	585						*HFC 218*
609	610			602	607	606	HFC 143a
				606			?
	647						?
				665			?
		678	683	680			HFC 1122
696			688	685			HFC 1234
	715			696			HFC 125
				700			CFC 22
721			712	707			CFC 114
727			717	713			CFC 114a
			721	717			?
740		720	728	723			HFC 1243
747				728			?
			733				HFC 245
749	759	725	735	729	726	729	HFC 134a
8802	885	883	884	884	890		HFC 134a + Cl
		989			989	990	?
1000		1001	999	999	1002	1000	HCFC 133
	1003						HFC 236
1013		1014	1011	1013	1016	1012	HFC 134
		1086			1088	1086	HCFC 123
		1093		1091	1095	1092	HCFC 133

Comparison of peaks (denoted by scan number) between unknown samples of HFC 134a based on mass spectral data (both for known and unknown spectra).

Unknown 1 HQ	Unknown 2 Q	Unknown 3 Red	Peak identity
182	182	183	Air
		395	CFC 115
613		604	HFC 143a
	613		HFC 32
676	674	682	HCFC 1122
680			HFC 1234
688			CFC 12B1
719		699	HFC 125
731	700	711	CFC 114
737	704	715	CFC 114a
		720	?
752	713		HFC 1243
764	717	732	HFC 134a
874	877	885	HFC 134A + Cl
1002	1002	1001	HCFC 133
1013	1015	1013	HFC 134
	1093		HCFC 133

Appendix M

Component scores for the first ten components calculated from the variance/co-variance matrix for the FID data of known samples.

Identity	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8	PC 9	PC 10
UK 1	-0.05	-0.04	0.28	1.16	0.93	-0.38	1.14	2.31	3.91	6.27
UK 2	-0.05	-0.03	0.19	0.37	1.08	-0.68	2.39	3.33	-1.12	0.56
UK 3	-0.05	-0.06	0.23	1.18	0.43	-0.11	0.95	0.52	3.46	5.97
UK 4	-0.05	-0.04	0.20	0.54	0.91	-0.35	1.03	2.58	1.19	-0.76
UK 5	-0.05	-0.05	0.21	0.95	0.48	-0.19	0.83	0.76	3.11	5.07
USA 1	0.19	-0.18	-0.15	-0.23	-0.25	-0.02	-0.28	-0.47	-0.06	-0.44
USA 2	0.18	-0.16	-0.02	-0.14	-0.59	0.05	-0.40	-0.77	-1.09	0.38
USA 3	0.19	-0.18	-0.13	-0.07	-0.62	0.40	-0.42	-0.15	4.58	3.58
USA 4	0.17	-0.14	-0.08	-0.05	-0.96	0.64	-0.33	0.23	3.51	2.54
USA 5	0.18	-0.15	-0.06	-0.10	-1.15	0.63	-0.29	0.01	2.92	3.42
Japan 1	-0.05	-0.06	-0.20	0.09	-0.58	-0.05	0.81	-0.60	-1.62	0.48
Japan 2	-0.05	-0.07	-0.18	0.00	-0.31	-0.28	-1.14	-0.03	-2.51	-0.69
Japan 3	-0.05	-0.06	-0.21	0.05	-0.50	-0.13	0.78	-0.55	-2.39	-0.65
Japan 4	-0.05	-0.07	-0.20	0.02	-0.31	-0.29	0.85	-0.76	-2.43	-0.60
Japan 5	-0.05	-0.07	-0.20	0.02	-0.39	-0.22	1.16	-0.85	-1.94	-0.38
Atochem 1	-0.05	-0.01	0.45	-0.41	0.20	-0.06	-0.88	0.35	-1.13	-2.56
Atochem 2	-0.06	-0.01	0.46	-0.30	0.03	-0.05	-1.03	-0.19	-1.14	-0.14
Atochem 3	-0.06	-0.01	0.42	-0.38	-0.02	-0.26	-0.98	-0.21	0.03	1.33
Atochem 4	-0.06	-0.02	0.44	-0.20	-0.06	-0.05	-1.07	-0.45	-0.69	0.52
Atochem 5	-0.06	-0.02	0.43	-0.27	-0.04	-0.05	-0.89	-0.36	-0.96	-0.32
Hoechst 1	0.09	0.49	0.13	-0.11	1.19	0.96	-0.78	0.61	5.82	-8.09
Hoechst 2	0.07	0.39	-0.01	-0.14	-0.66	-0.63	-1.00	-0.30	0.71	-3.61
Hoechst 3	0.07	0.43	0.03	-0.08	2.17	2.61	-0.88	0.00	2.21	-5.17
Hoechst 4	0.07	0.40	-0.02	-0.18	-0.67	-0.62	-0.83	-0.33	0.71	-4.02
Hoechst 5	0.07	0.43	0.03	-0.06	1.66	2.35	-0.89	-0.22	1.71	-3.91
Showa 1	-0.05	-0.07	-0.22	-0.33	-0.11	-0.42	-1.07	0.47	-2.40	-0.81
Showa 2	-0.05	-0.07	-0.22	-0.21	-0.16	-0.35	-1.07	0.13	-1.77	0.03
Showa 3	-0.05	-0.07	-0.23	-0.18	-0.21	-0.36	-1.15	0.00	-1.80	0.77
Showa 4	-0.05	-0.07	-0.21	-0.19	-0.18	-0.33	-1.07	0.03	-1.66	0.30
Showa 5	-0.05	-0.07	-0.21	-0.18	-0.17	-0.33	-1.08	0.04	-1.51	0.41
Ashai 1	-0.05	-0.07	-0.19	-0.27	-0.15	-0.32	-0.84	0.05	-1.63	-0.22
Ashai 2	-0.05	-0.07	-0.19	-0.05	-0.26	-0.27	1.86	-1.22	-1.60	0.01
Ashai 3	-0.05	-0.07	-0.20	-0.07	-0.24	-0.29	2.80	-1.51	-1.53	0.10
Ashai 4	-0.05	-0.07	-0.19	-0.07	-0.25	-0.28	2.09	-1.30	-1.49	0.23
Ashai 5	-0.05	-0.07	-0.19	-0.09	-0.24	-0.28	1.68	-1.16	-1.41	0.42

Appendix N

Component scores for the first ten components calculated from the correlation matrix for the FID data of known samples.

Identity	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8	PC 9	PC 10
UK 1	-0.17	-0.87	2.74	1.05	-1.26	0.10	-0.37	-0.65	0.41	0.51
UK 2	-0.15	-0.70	1.32	0.25	-0.52	-1.49	-1.75	-0.40	-3.47	0.41
UK 3	-0.28	-0.81	2.26	1.12	-0.59	0.82	0.42	-0.11	2.05	0.21
UK 4	-0.20	-0.70	1.53	0.38	0.07	-0.07	0.05	1.68	-2.44	-0.81
UK 5	-0.26	-0.76	1.96	0.90	-0.42	0.72	0.39	0.24	1.33	-0.05
USA 1	-0.93	2.26	0.11	-0.16	-0.24	-1.12	-2.15	2.06	0.91	0.14
USA 2	-0.88	2.06	0.21	-0.19	-0.08	-0.50	-1.96	1.86	1.34	-0.19
USA 3	-0.99	2.38	0.37	0.04	-0.29	0.03	1.16	-1.93	-0.46	0.52
USA 4	-0.89	2.04	0.43	0.06	0.41	0.79	1.48	-0.49	-0.86	-0.18
USA 5	-0.95	2.28	0.36	-0.03	0.33	0.78	1.33	-1.27	-0.89	-0.03
Japan 1	-0.41	-0.53	-0.51	0.75	0.69	-0.14	0.15	-1.25	0.36	-0.39
Japan 2	-0.32	-0.39	-0.45	0.00	-0.91	-0.40	-0.43	-0.82	0.27	-4.08
Japan 3	-0.42	-0.58	-0.83	0.81	0.64	-0.85	-0.44	-1.44	0.77	0.61
Japan 4	-0.41	-0.54	-0.48	0.57	0.68	-0.21	-0.34	-0.23	-0.02	-1.15
Japan 5	-0.48	-0.58	-0.36	0.78	1.95	1.03	0.93	1.14	-0.96	-0.85
Atochem 1	0.05	-0.51	0.14	-2.47	0.08	-0.79	-0.28	-0.55	-0.57	0.20
Atochem 2	0.02	-0.55	0.31	-2.37	0.10	-0.25	-0.11	-0.69	0.45	0.28
Atochem 3	-0.01	-0.54	0.26	-2.29	-0.02	-0.12	-0.23	-0.91	0.77	0.62
Atochem 4	-0.02	-0.58	0.47	-2.16	0.39	0.38	0.28	-0.07	0.75	0.14
Atochem 5	-0.07	-0.59	0.49	-2.03	1.37	1.24	0.95	1.85	-0.24	-0.28
Hoechst 1	2.72	0.68	0.15	0.21	-0.22	-0.87	-0.43	-0.74	-0.34	0.96
Hoechst 2	2.10	0.50	-0.48	0.22	-0.33	2.62	-1.62	-0.56	-0.25	0.58
Hoechst 3	2.36	0.55	0.09	0.39	0.06	-2.03	2.02	1.21	0.17	0.00
Hoechst 4	2.10	0.55	-0.44	0.30	0.78	2.43	-1.64	-0.34	-0.21	-1.03
Hoechst 5	2.30	0.53	0.12	0.36	0.30	-1.63	1.49	0.45	0.80	-0.69
Showa 1	-0.26	-0.35	-1.48	-0.08	-2.28	-0.30	-0.04	-0.17	-0.65	0.72
Showa 2	-0.31	-0.42	-1.18	0.16	-1.47	0.69	0.73	1.08	-0.36	0.65
Showa 3	-0.30	-0.40	-1.17	0.15	-1.81	0.46	0.47	0.38	0.13	0.37
Showa 4	-0.33	-0.43	-1.12	0.21	-1.22	0.70	0.76	1.02	-0.11	0.41
Showa 5	-0.32	-0.44	-1.14	0.24	-1.34	0.56	0.72	0.76	0.04	0.74
Ashai 1	-0.38	-0.36	-0.99	-0.01	-0.38	-0.85	-0.19	-0.68	0.37	-2.03
Ashai 2	-0.46	-0.52	-0.70	0.65	1.04	-0.65	-0.48	-0.52	0.40	0.64
Ashai 3	-0.50	-0.59	-0.64	0.86	1.90	-0.33	-0.33	0.22	-0.05	1.61
Ashai 4	-0.48	-0.55	-0.61	0.71	1.50	-0.23	-0.22	0.22	0.11	0.70
Ashai 5	-0.46	-0.53	-0.73	0.64	1.11	-0.54	-0.32	-0.36	0.45	0.73

Appendix O

Component scores for the first ten components calculated from the correlation matrix for the ECD data of known samples.

Identity	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8	PC 9	PC 10
UK 1	0.34	-0.10	0.73	1.96	1.53	-0.85	0.31	0.18	-0.17	-2.34
UK 2	0.10	-0.11	0.74	3.06	2.30	-1.80	-0.48	-1.38	-0.14	4.35
UK 3	0.05	-0.28	0.05	0.74	0.82	-1.31	0.25	-0.61	-0.25	-2.15
UK 4	0.04	-0.31	0.03	0.65	0.88	-1.37	0.21	-0.59	-0.36	-2.04
UK 5	0.10	-0.25	0.21	1.21	1.14	-1.21	0.10	-0.69	-0.28	-1.99
USA 1	-0.65	2.52	0.10	-0.53	0.19	0.01	-0.11	-0.14	-0.14	0.00
USA 2	-0.63	2.51	0.10	-0.56	0.21	-0.03	-0.13	-0.09	-0.15	0.02
USA 3	-0.62	2.52	0.14	-0.42	0.16	0.08	-0.06	-0.12	-0.15	0.03
USA 4	-0.60	2.61	0.21	-0.33	0.21	0.18	-0.09	-0.05	-0.05	-0.02
USA 5	-0.67	2.84	0.34	0.01	0.21	0.49	0.13	0.01	0.10	-0.02
Japan 1	-0.45	-0.45	-0.74	-1.16	-0.40	-1.64	-0.18	0.31	0.04	0.41
Japan 2	-0.39	-0.23	-0.23	0.21	-0.46	-0.94	0.51	2.19	1.51	0.42
Japan 3	-0.39	-0.22	-0.22	0.35	-0.52	-0.84	0.58	2.48	1.79	0.23
Japan 4	-0.38	-0.20	-0.16	0.41	-0.46	-0.71	0.51	2.10	1.53	0.15
Japan 5	-0.36	-0.14	0.01	0.75	-0.50	-0.56	0.62	2.12	1.63	0.02
Atochem 1	-0.27	-0.84	1.76	-1.56	0.88	0.24	0.16	0.06	0.05	0.24
Atochem 2	-0.27	-0.82	1.77	-1.20	0.85	0.63	0.28	0.12	0.05	0.16
Atochem 3	-0.26	-0.75	1.57	-0.97	0.86	0.55	0.64	0.01	0.01	0.11
Atochem 4	-0.25	-0.90	2.10	-1.34	1.05	0.86	0.22	0.12	0.09	0.09
Atochem 5	-0.27	-0.90	2.12	-1.32	1.04	0.92	0.31	0.17	0.06	0.31
Hoechst 1	2.45	0.22	-0.22	-0.22	-0.32	0.12	-0.06	-0.50	0.86	-0.02
Hoechst 2	2.26	0.17	-0.54	-0.85	-0.50	0.23	0.63	-3.14	4.26	0.29
Hoechst 3	2.72	0.26	-0.15	0.00	0.07	0.36	-0.10	0.61	-0.82	-0.57
Hoechst 4	2.82	0.31	-0.34	-0.32	-0.15	0.14	0.33	1.92	-2.27	1.69
Hoechst 5	2.97	0.29	-0.23	-0.21	0.06	0.24	0.02	0.78	-1.18	-0.52
Showa 1	-0.22	-0.28	0.60	1.06	-1.52	0.69	-1.30	-0.21	-0.16	0.10
Showa 2	-0.20	-0.34	0.58	0.74	-1.05	0.52	-2.09	-0.08	0.10	-0.14
Showa 3	-0.16	-0.35	0.59	0.70	-1.15	0.60	-2.13	-0.11	0.04	-0.13
Showa 4	-0.11	-0.34	0.72	0.89	-1.33	0.85	-2.54	-0.16	0.02	-0.24
Showa 5	-0.16	-0.41	0.36	-0.01	-1.15	0.18	-2.44	0.01	0.08	-0.29
Ashai 1	-0.43	-0.37	-0.16	0.14	-1.10	-0.32	0.71	-0.71	-0.67	0.07
Ashai 2	-0.48	-0.33	0.08	0.61	-1.66	0.61	2.07	-0.74	-0.78	0.01
Ashai 3	-0.49	-0.31	0.16	0.87	-1.55	0.56	1.88	-0.82	-0.87	0.42
Ashai 4	-0.47	-0.32	0.06	0.62	-1.46	0.35	1.62	-0.71	-0.74	0.13
Ashai 5	-0.48	-0.33	0.15	0.64	-1.61	0.54	2.00	-0.73	-0.79	0.14
Air 1	-0.55	-0.46	-1.66	0.19	0.84	1.04	-0.17	-0.02	0.21	-0.03
Air 2	-0.57	-0.51	-1.83	0.35	1.19	1.82	-0.04	0.11	0.25	0.01

Air 3	-0.57	-0.59	-2.13	0.38	1.60	2.36	-0.05	0.25	-0.05	0.05
Air 4	-0.57	-0.55	-1.97	0.31	1.36	2.02	-0.06	0.16	0.06	0.02
Helium 1	-0.50	-0.58	-1.26	-1.53	-0.11	-1.39	-0.55	-0.58	-0.65	0.36
Helium 2	-0.48	-0.57	-1.19	-1.49	-0.13	-1.42	-0.52	-0.52	-0.70	0.29
Helium 3	-0.48	-0.56	-1.17	-1.45	-0.14	-1.39	-0.52	-0.53	-0.68	0.25
Helium 4	-0.46	-0.54	-1.06	-1.39	-0.19	-1.41	-0.48	-0.48	-0.68	0.15

Appendix P

First ten PC scores for joint analysis of FID and ECD data for known samples of HFC 134a

Identity	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8	PC 9	PC 10
UK 1	0.06	-0.56	1.32	2.67	0.87	0.61	-0.02	-0.45	1.39	1.03
UK 2	-0.06	-0.56	1.04	2.10	1.52	0.65	-1.72	3.03	-2.74	-1.52
UK 3	-0.11	-0.58	0.88	1.89	-0.28	-1.01	1.65	-1.10	0.80	-0.27
UK 4	-0.08	-0.56	0.80	1.29	-0.33	-0.89	0.86	-0.58	-0.05	-0.61
UK 5	-0.09	-0.57	0.92	1.84	0.15	-0.80	1.47	-0.73	0.24	-0.28
USA 1	-0.83	2.25	0.25	-0.21	-0.03	-0.22	0.03	0.03	-0.46	-0.32
USA 2	-0.80	2.13	0.31	-0.18	-0.10	-0.32	0.05	-0.04	-0.22	-0.41
USA 3	-0.84	2.31	0.30	-0.03	0.06	-0.17	0.19	-0.12	-0.17	0.08
USA 4	-0.80	2.16	0.33	0.10	0.03	0.11	0.03	-0.18	0.37	0.09
USA 5	-0.88	2.37	0.29	0.15	0.19	0.67	-0.08	0.20	0.40	0.52
Japan 1	-0.42	-0.53	-0.44	-0.52	-1.50	-3.61	0.31	-0.22	-1.66	-0.92
Japan 2	-0.42	-0.38	-0.52	0.08	-0.57	-1.17	-1.90	0.44	0.40	0.85
Japan 3	-0.47	-0.48	-0.94	0.25	-0.93	-0.93	-1.90	0.31	0.45	1.17
Japan 4	-0.46	-0.45	-0.70	0.33	-0.76	-0.64	-1.58	0.27	0.60	0.86
Japan 5	-0.49	-0.43	-0.75	0.71	-0.79	0.43	-2.02	0.12	1.90	0.91
Atochem 1	-0.13	-0.73	1.74	-1.68	-0.16	-0.22	-0.26	0.19	-0.57	-0.05
Atochem 2	-0.15	-0.72	1.68	-1.42	-0.17	0.41	-0.19	0.07	0.01	0.15
Atochem 3	-0.17	-0.69	1.54	-1.22	-0.24	0.52	-0.07	0.24	-0.11	0.20
Atochem 4	-0.17	-0.77	1.83	-1.40	-0.15	0.58	-0.04	-0.20	0.43	0.20
Atochem 5	-0.20	-0.77	1.81	-1.38	-0.35	0.77	-0.16	-0.32	0.71	-0.01
Hoechst 1	2.46	0.49	-0.09	-0.15	0.04	-0.21	0.42	0.68	-1.02	1.36
Hoechst 2	2.06	0.40	-0.45	-0.52	-0.45	-0.82	2.20	3.77	2.00	0.72
Hoechst 3	2.39	0.46	-0.18	0.12	0.19	0.41	-0.46	-1.55	-1.41	1.46
Hoechst 4	2.31	0.52	-0.46	-0.16	-0.57	0.54	-1.43	-0.59	1.86	-4.34
Hoechst 5	2.48	0.48	-0.17	0.14	-0.10	0.28	-0.62	-1.82	-1.22	0.87
Showa 1	-0.33	-0.44	-1.17	-0.65	1.95	0.39	0.17	0.27	-0.38	0.20
Showa 2	-0.33	-0.47	-0.93	-0.53	1.95	-0.18	0.17	-0.20	0.36	-0.08
Showa 3	-0.30	-0.46	-0.96	-0.57	2.02	-0.19	0.28	-0.28	0.28	-0.07
Showa 4	-0.30	-0.47	-1.02	-0.53	2.26	0.11	0.34	-0.52	0.53	-0.09
Showa 5	-0.29	-0.48	-0.93	-0.78	1.54	-1.07	0.30	-0.73	0.45	-0.44
Ashai 1	-0.45	-0.43	-0.85	-0.43	-0.49	-0.32	0.49	-0.02	-1.08	-0.63
Ashai 2	-0.55	-0.50	-1.14	0.08	-1.25	1.52	1.07	0.01	-0.63	0.07
Ashai 3	-0.57	-0.53	-1.15	0.28	-1.23	1.77	0.86	0.08	-0.67	-0.31
Ashai 4	-0.54	-0.50	-1.06	0.20	-1.15	1.36	0.77	-0.07	-0.37	-0.33
Ashai 5	-0.54	-0.50	-1.12	0.11	-1.18	1.64	0.80	0.00	-0.42	-0.05

Appendix Q

Table of variations for the first twenty PCs extracted form the correlation coefficients matrix of all samples of HFC 134a analysed by FID.

	Initial Eigenvalues			Considered components		
Component	Value	% initial of variance	Cumulative percentage	Total	% of Variance	Cumulative %
1	11.66	30.69	30.7	11.66	30.69	30.7
2	7.66	20.16	50.8	7.66	20.16	50.8
3	3.80	9.99	60.8	3.80	9.99	60.8
4	2.94	7.73	68.6	2.94	7.73	68.6
5	2.50	6.57	75.1	2.50	6.57	75.1
6	2.05	5.39	80.5	2.05	5.39	80.5
7	1.72	4.52	85.0	1.72	4.52	85.0
8	1.52	4.00	89.0	1.52	4.00	89.0
9	1.02	2.68	91.7	1.02	2.68	91.7
10	0.77	2.03	93.8	0.77	2.03	93.8
11	0.60	1.59	95.3			
12	0.53	1.38	96.7			
13	0.36	0.96	97.7			
14	0.20	0.52	98.2			
15	0.13	0.33	98.5			
16	0.12	0.32	98.9			
17	0.10	0.25	99.1			
18	0.07	0.19	99.3			
19	0.06	0.15	99.4			
20	0.05	0.12	99.6			

Appendix R

First ten PC scores for all samples of HFC 134a analysed by GC FID.

Identity	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8	PC 9	PC 10
UK 1	-0.15	-0.58	0.88	0.30	-0.36	3.27	-0.10	-0.57	-0.25	2.70
UK 2	-0.13	-0.58	0.06	0.03	0.07	2.18	-0.02	-0.50	0.73	0.93
UK 3	-0.22	-0.54	0.89	0.07	-0.75	3.16	-0.03	-0.21	-0.30	1.93
UK 4	-0.15	-0.56	0.24	0.27	-0.38	2.69	0.15	-0.16	0.67	1.06
UK 5	-0.20	-0.53	0.71	0.09	-0.73	2.92	-0.02	-0.11	-0.13	1.66
USA 1	-0.71	2.71	-0.07	0.69	0.35	-0.41	0.47	1.06	3.49	1.89
USA 2	-0.67	2.44	0.05	0.63	0.21	-0.42	0.44	0.87	2.56	0.93
USA 3	-0.75	2.95	-0.03	0.72	0.32	-0.16	0.31	0.59	1.51	1.31
USA 4	-0.65	2.59	0.04	0.70	0.06	0.31	0.34	0.77	1.85	0.46
USA 5	-0.71	2.82	0.03	0.71	0.20	-0.10	0.34	0.70	1.71	0.09
Japan 1	-0.34	-0.31	-0.20	-1.05	-0.34	0.95	-0.25	0.72	0.03	-1.27
Japan 2	-0.33	-0.29	-0.25	-0.74	-0.29	-0.44	-0.31	-0.32	-0.25	0.00
Japan 3	-0.33	-0.38	-0.34	-1.16	-0.04	0.71	-0.33	0.93	-0.01	-0.98
Japan 4	-0.37	-0.39	-0.22	-1.00	-0.47	0.93	-0.11	0.63	0.45	-1.29
Japan 5	-0.41	-0.38	-0.11	-0.94	-0.72	1.78	0.14	1.12	0.91	-2.57
Atochem 1	0.04	-0.70	-0.87	1.39	-0.35	-0.38	0.09	-0.05	-0.05	0.42
Atochem 2	0.00	-0.71	-0.71	1.38	-0.40	-0.29	0.09	-0.05	-0.16	0.25
Atochem 3	-0.04	-0.69	-0.70	1.27	-0.48	-0.26	0.08	-0.01	-0.23	0.38
Atochem 4	-0.04	-0.71	-0.57	1.35	-0.55	0.07	0.14	0.08	-0.05	-0.03
Atochem 5	-0.07	-0.72	-0.50	1.36	-0.79	0.69	0.38	0.41	0.56	-1.00
Hoechst 1	4.51	0.80	0.26	0.13	0.20	0.32	-0.68	-0.04	-0.15	1.17
Hoechst 2	3.55	0.66	0.36	-0.78	-0.02	-0.83	-0.97	-0.06	0.37	-2.39
Hoechst 3	4.07	0.71	0.30	0.04	-0.10	0.73	-1.21	0.07	0.23	1.84
Hoechst 4	3.54	0.69	0.45	-0.59	0.03	-0.36	-0.34	0.16	0.34	-3.04
Hoechst 5	3.96	0.70	0.39	0.07	-0.06	0.57	-1.02	0.12	-0.01	1.08
Showa 1	-0.27	-0.32	-0.78	-1.44	-0.27	-1.42	-0.79	-0.33	0.05	1.09
Showa 2	-0.29	-0.34	-0.56	-1.34	-0.53	-0.67	-0.61	0.03	0.31	0.47
Showa 3	-0.30	-0.32	-0.53	-1.37	-0.49	-0.90	-0.69	-0.14	0.08	0.72
Showa 4	-0.30	-0.34	-0.50	-1.27	-0.52	-0.52	-0.56	0.11	0.27	0.33
Showa 5	-0.30	-0.34	-0.52	-1.30	-0.47	-0.57	-0.60	0.12	0.18	0.51
Ashai 1	-0.34	-0.24	-0.50	-0.78	-0.16	-0.37	-0.32	0.20	-0.07	0.10
Ashai 2	-0.40	-0.34	-0.32	-1.09	-0.56	1.28	-0.11	0.94	0.63	-1.18
Ashai 3	-0.42	-0.39	-0.28	-1.16	-0.76	2.01	0.05	1.36	1.10	-1.96
Ashai 4	-0.41	-0.37	-0.26	-1.07	-0.72	1.65	0.00	1.11	0.90	-1.63
Ashai 5	-0.40	-0.34	-0.32	-1.06	-0.56	1.27	-0.10	1.01	0.63	-1.19
HQ 1	0.03	-0.85	-1.34	-0.34	4.56	0.60	-0.21	3.03	-1.37	1.74
HQ 2	-0.06	-0.72	-1.18	-0.62	3.88	-0.51	-0.22	2.15	-1.26	0.74
HQ 3	-0.12	-0.78	-1.11	-0.61	3.69	0.32	-0.06	2.92	-0.66	-0.67
Q 1	-0.26	-0.83	0.42	0.20	3.11	0.60	-0.07	-3.83	2.13	-1.17
Q 2	-0.23	-0.74	0.47	-0.09	2.78	0.47	0.07	-3.52	1.85	-1.03

Q 3	-0.25	-0.76	0.40	-0.01	2.97	0.27	-0.11	-3.75	1.72	-0.66
Red 1	0.82	-0.18	0.15	-1.46	-0.34	-0.08	5.28	-0.11	-0.22	-0.43
Red 2	0.80	-0.13	-0.02	-1.35	0.02	-1.17	4.74	-0.56	-0.74	0.62
Red 3	0.92	-0.15	0.05	-1.07	-0.21	-0.13	5.09	-0.28	-0.48	1.01
QF1	-0.49	1.55	-0.36	0.01	-0.02	0.19	-0.18	-0.55	-1.55	-0.62
QF2	-0.50	1.55	-0.28	-0.02	0.21	-0.06	-0.32	-0.84	-1.91	-0.19
QF3	-0.47	1.61	-0.46	0.06	-0.02	0.37	-0.11	-0.48	-1.20	-1.03
QF4	-0.52	2.15	-0.49	0.26	0.28	0.13	-0.14	-0.71	-1.60	-0.29
QF5	-0.52	2.09	-0.44	0.29	0.05	0.60	-0.01	-0.54	-1.55	-1.14
QF6	-0.53	1.76	-0.23	0.11	0.12	0.32	-0.13	-0.70	-1.78	-0.80
QF7	-0.54	1.83	-0.39	0.15	0.07	0.29	-0.07	-0.51	-1.20	-0.69
QF8	-0.51	1.63	-0.32	0.02	0.19	0.00	-0.23	-0.74	-1.62	-0.24
QF9	-0.52	1.76	-0.34	0.12	0.03	0.44	-0.12	-0.64	-1.83	-1.15
QF10	-0.48	1.57	-0.47	-0.02	0.16	-0.10	-0.28	-0.76	-1.78	-0.27
QH 1	-0.28	-0.31	-0.57	-1.21	-0.59	-0.94	-0.60	-0.42	0.31	0.60
QH 2	-0.28	-0.31	-0.60	-1.26	-0.60	-0.97	-0.61	-0.42	0.37	0.57
QH 3	-0.28	-0.31	-0.56	-1.21	-0.58	-0.93	-0.59	-0.42	0.29	0.60
QH 4	-0.28	-0.30	-0.56	-1.27	-0.56	-1.01	-0.64	-0.47	0.21	0.74
QH 5	-0.27	-0.31	-0.58	-1.27	-0.58	-1.02	-0.65	-0.47	0.23	0.85
QH 6	-0.28	-0.33	-0.54	-1.28	-0.73	-0.78	-0.56	-0.33	0.56	0.34
QH 7	-0.27	-0.33	-0.61	-1.37	-0.75	-0.90	-0.64	-0.40	0.65	0.54
QH 8	-0.27	-0.32	-0.58	-1.37	-0.69	-0.94	-0.64	-0.43	0.50	0.65
QH 9	-0.28	-0.32	-0.57	-1.26	-0.65	-0.87	-0.60	-0.40	0.42	0.57
QH 10	-0.28	-0.31	-0.58	-1.23	-0.58	-0.96	-0.61	-0.43	0.29	0.65
97Q1	-0.50	-0.16	4.08	0.77	0.41	-2.26	-0.13	2.39	1.70	-0.93
97Q2	-0.38	-0.28	1.98	0.14	0.10	-0.88	-0.25	0.42	-0.17	0.15
97Q3	-0.38	-0.29	1.93	0.14	0.22	-0.89	-0.26	0.34	-0.08	0.06
97Q4	-0.36	-0.35	1.73	0.09	0.19	-0.55	-0.33	-0.14	-0.57	0.43
97Q5	-0.43	-0.41	2.18	-0.05	-0.27	0.24	-0.42	-0.22	-1.64	0.52
97Q6	-0.46	-0.35	2.41	0.13	-0.23	0.25	-0.16	0.29	-0.53	-0.37
97Q7	-0.42	-0.34	2.13	0.17	-0.11	-0.07	-0.13	0.40	-0.10	-0.49
97Q8	-0.31	-0.58	1.65	1.25	-0.42	-0.72	-0.17	0.37	-0.53	-0.51
97Q9	-0.39	-0.37	1.84	-0.07	0.09	-0.65	-0.43	-0.23	-1.32	0.68
97Q10	-0.18	-0.34	2.41	0.22	0.59	-1.43	0.75	0.73	0.13	-0.30
97Q11	-0.39	-0.31	1.75	-0.04	0.00	-0.42	-0.36	-0.17	-1.01	0.63
97Q12	-0.41	-0.41	1.86	-0.09	-0.32	0.50	-0.32	-0.38	-1.56	0.33
97Q13	-0.39	-0.41	1.55	-0.05	-0.09	-0.10	-0.26	-0.18	-0.95	-0.19
98Q1	0.04	-0.67	-0.76	1.24	-0.38	-0.39	0.12	0.02	-0.03	0.00
98Q2	0.16	-0.80	-0.81	1.88	-0.43	-0.02	0.15	0.21	0.24	-0.42
98Q3	0.11	-0.77	-0.76	1.68	-0.47	-0.15	0.21	0.21	0.22	-0.68
98Q4	0.11	-0.72	-0.83	1.51	-0.31	-0.30	0.01	0.03	0.11	-0.19
98Q5	0.10	-0.70	-0.77	1.43	-0.41	-0.20	0.09	0.10	0.08	-0.11
98Q6	0.14	-0.76	-0.85	1.68	-0.29	-0.40	0.06	0.03	0.15	-0.52
98Q7	0.03	-0.66	-0.91	1.21	-0.12	-0.96	0.00	-0.35	-0.44	0.68

98Q8	0.04	-0.69	-0.82	1.34	-0.31	-0.74	0.11	-0.13	-0.28	0.14
98Q9	0.06	-0.75	-0.85	1.58	-0.30	-0.60	0.21	-0.08	-0.18	0.04
98Q10	0.07	-0.79	-0.87	1.73	-0.28	-0.48	0.21	-0.03	-0.08	-0.10
98Q11	0.11	-0.77	-0.72	1.72	-0.61	-0.09	0.33	0.27	0.14	-0.54
98Q12	0.07	-0.75	-0.85	1.57	-0.26	-0.50	0.10	-0.03	-0.03	-0.21
98Q13	0.09	-0.73	-0.76	1.53	-0.45	-0.38	0.19	0.13	0.11	-0.67
98Q14	0.07	-0.69	-0.78	1.33	-0.36	-0.45	0.07	-0.01	-0.05	-0.05

Appendix S

Agglomeration schedule for PC scores extracted from correlation coefficients matrix all samples of HFC 134a analysed by GC FID.

	Clusters Combined		Coefficients	Stage Cluster First Appears		Next Stage
Stage	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	55	57	0.008	0	0	2
2	55	64	0.020	1	0	15
3	32	35	0.032	0	0	49
4	56	59	0.045	0	0	6
5	86	89	0.067	0	0	19
6	56	58	0.089	4	0	15
7	27	28	0.121	0	0	17
8	29	30	0.153	0	0	17
9	85	91	0.186	0	0	20
10	61	62	0.222	0	0	29
11	60	63	0.264	0	0	29
12	83	87	0.307	0	0	43
13	51	53	0.358	0	0	31
14	46	52	0.408	0	0	36
15	55	56	0.461	2	6	55
16	33	34	0.514	0	0	46
17	27	29	0.569	7	8	55
18	82	90	0.626	0	0	50
19	81	86	0.682	0	5	43
20	16	85	0.742	0	9	44
21	66	67	0.806	0	0	66
22	45	47	0.871	0	0	51
23	73	75	0.939	0	0	40
24	17	18	1.011	0	0	26
25	80	88	1.088	0	0	42
26	17	78	1.168	24	0	44
27	11	14	1.248	0	0	49
28	23	25	1.328	0	0	69
29	60	61	1.413	11	10	61
30	3	5	1.505	0	0	62
31	50	51	1.597	0	13	64
32	8	10	1.691	0	0	39
33	22	24	1.796	0	0	78
34	37	38	1.905	0	0	71
35	40	41	2.015	0	0	54
36	46	54	2.127	14	0	51
37	7	9	2.241	0	0	65

38	48	49	2.361	0	0	73
39	6	8	2.482	0	32	65
40	68	73	2.607	0	23	52
41	19	20	2.733	0	0	70
42	79	80	2.870	0	25	68
43	81	83	3.009	19	12	68
44	16	17	3.150	20	26	50
45	69	71	3.293	0	0	48
46	15	33	3.436	0	16	63
47	12	31	3.579	0	0	59
48	69	70	3.732	45	0	72
49	11	32	3.895	27	3	63
50	16	82	4.063	44	18	57
51	45	46	4.249	22	36	64
52	68	77	4.438	40	0	58
53	43	44	4.630	0	0	56
54	39	40	4.827	0	35	82
55	27	55	5.044	17	15	61
56	42	43	5.263	0	53	83
57	16	84	5.516	50	0	70
58	68	76	5.781	52	0	66
59	12	13	6.049	47	0	74
60	2	4	6.319	0	0	76
61	27	60	6.594	55	29	67
62	1	3	6.870	0	30	76
63	11	15	7.162	49	46	74
64	45	50	7.455	51	31	73
65	6	7	7.764	39	37	84
66	66	68	8.079	21	58	75
67	26	27	8.421	0	61	80
68	79	81	8.774	42	43	77
69	21	23	9.142	0	28	78
70	16	19	9.592	57	41	77
71	36	37	10.13	0	34	82
72	69	74	10.72	48	0	75
73	45	48	11.34	64	38	84
74	11	12	11.98	63	59	80
75	66	69	12.68	66	72	79
76	1	2	13.42	62	60	85
77	16	79	14.28	70	68	89
78	21	22	15.27	69	33	86
79	66	72	16.32	75	0	81
80	11	26	17.93	74	67	83
81	65	66	19.79	0	79	85

82	36	39	22.44	71	54	86
83	11	42	25.31	80	56	90
84	6	45	28.45	65	73	88
85	1	65	32.28	76	81	87
86	21	36	45.41	78	82	87
87	1	21	60.86	85	86	88
88	1	6	80.21	87	84	89
89	1	16	102.6	88	77	90
90	1	11	127.4	89	83	0

Appendix T

Table of variations for the first twenty PCs extracted form the correlation coefficients matrix of all samples of HFC 134a analysed by ECD.

	Initial Eigenvalues			Considered components		
Component	Value	% of initial variance	Cumulative percentage	Value	% of initial variance	Cumulative percentage
1	12.0	26.0	26.0	12.0	26.0	26.0
2	6.9	15.1	41.0	6.9	15.1	41.0
3	5.3	11.6	52.6	5.3	11.6	52.6
4	3.6	7.9	60.5	3.6	7.9	60.5
5	3.2	7.0	67.6	3.2	7.0	67.6
6	2.4	5.2	72.7	2.4	5.2	72.7
7	2.1	4.5	77.2	2.1	4.5	77.2
8	1.7	3.7	80.9	1.7	3.7	80.9
9	1.5	3.2	84.1	1.5	3.2	84.1
10	1.4	3.1	87.3	1.4	3.1	87.3
11	1.1	2.4	89.7			
12	0.9	1.9	91.6			
13	0.7	1.5	93.1			
14	0.6	1.2	94.3			
15	0.5	1.0	95.3			
16	0.4	0.9	96.3			
17	0.3	0.7	97.0			
18	0.3	0.6	97.6			
19	0.2	0.5	98.1			
20	0.2	0.4	98.5			

Appendix U

First ten PC scores for all samples of HFC 134a analysed by GC ECD.

Identity	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8	PC 9	PC 10
UK 1	0.31	-0.33	0.63	-1.63	0.31	0.72	0.48	-0.62	2.17	-0.14
UK 2	0.10	-0.56	0.72	-2.21	0.37	1.11	0.63	2.84	1.66	-0.02
UK 3	0.09	-0.50	0.35	-0.90	0.34	-0.27	0.15	-0.44	-0.11	-0.06
UK 4	0.09	-0.52	0.33	-0.94	0.33	-0.27	0.12	-0.47	-0.21	-0.07
UK 5	0.11	-0.48	0.50	-1.14	0.28	-0.12	0.26	-0.32	0.39	0.01
USA 1	-1.25	2.48	0.16	-0.22	0.28	-0.28	-1.18	0.26	0.58	-2.31
USA 2	-1.21	2.38	0.17	-0.23	0.22	-0.22	-1.06	0.19	0.33	-1.83
USA 3	-1.21	2.36	0.18	-0.19	0.22	-0.18	-0.82	0.21	0.45	-1.72
USA 4	-1.22	2.41	0.23	-0.23	0.20	-0.09	-0.84	0.27	0.72	-1.74
USA 5	-1.40	2.87	0.31	-0.31	0.35	0.02	-0.81	0.47	1.81	-3.04
Japan 1	-0.29	-0.55	-0.25	0.04	-0.22	-0.65	-0.21	-0.41	-2.54	-1.39
Japan 2	-0.35	-0.43	0.00	0.05	0.27	-0.19	1.07	0.00	-0.48	-2.82
Japan 3	-0.38	-0.42	0.02	0.22	0.26	-0.26	1.17	-0.05	-0.45	-3.17
Japan 4	-0.37	-0.40	0.04	0.19	0.21	-0.20	1.08	-0.05	-0.28	-2.69
Japan 5	-0.37	-0.38	0.04	0.13	0.28	-0.05	1.21	0.07	0.20	-2.48
Atochem 1	-0.07	-0.68	-0.45	0.04	-1.26	0.69	-1.25	-0.92	-0.22	-0.04
Atochem 2	-0.10	-0.61	-0.38	-0.01	-0.87	0.82	-0.78	-0.78	0.32	0.03
Atochem 3	-0.11	-0.65	-0.37	0.01	-0.89	0.96	-0.79	-0.66	0.59	-0.12
Atochem 4	-0.05	-0.67	-0.44	0.03	-1.06	1.00	-0.98	-0.93	0.75	0.16
Atochem 5	-0.08	-0.64	-0.40	-0.02	-1.04	1.07	-0.96	-0.76	0.68	0.12
Hoechst 1	2.70	0.91	2.65	0.73	-0.02	0.27	-0.04	-0.79	0.07	-0.05
Hoechst 2	2.61	1.03	1.57	1.05	0.27	0.36	0.08	-0.69	-0.52	-0.11
Hoechst 3	2.83	0.96	2.91	0.85	0.08	0.35	0.00	-0.73	0.36	0.04
Hoechst 4	2.94	1.06	3.58	1.21	0.12	1.29	-0.08	-0.52	-1.32	-0.42
Hoechst 5	2.97	0.99	3.01	0.83	0.09	0.39	-0.08	-0.84	0.06	0.03
Showa 1	-0.22	-0.33	0.28	0.99	-0.56	-1.22	1.74	0.89	1.20	-0.01
Showa 2	-0.15	-0.44	0.24	0.86	-0.51	-1.40	0.52	0.84	0.97	0.41
Showa 3	-0.09	-0.46	0.29	1.01	-0.59	-1.56	0.41	0.98	1.00	0.44
Showa 4	-0.05	-0.47	0.38	1.26	-0.75	-1.91	0.40	1.23	1.41	0.63
Showa 5	-0.04	-0.52	0.21	1.17	-0.65	-1.99	-0.32	0.99	0.22	0.40
Ashai 1	-0.36	-0.50	-0.19	0.48	-0.15	-0.22	1.14	-0.20	-1.03	-0.37
Ashai 2	-0.51	-0.33	-0.13	0.78	-0.04	0.48	2.79	0.13	0.29	-0.81
Ashai 3	-0.50	-0.36	-0.13	0.68	0.00	0.33	2.80	0.16	0.19	-0.73
Ashai 4	-0.46	-0.40	-0.15	0.67	-0.01	0.22	2.31	0.07	-0.05	-0.63
Ashai 5	-0.49	-0.36	-0.15	0.71	-0.05	0.43	2.66	0.13	0.18	-0.76
Air 1	-0.53	-0.78	-0.22	1.14	3.07	0.47	-0.39	-0.03	0.08	0.22
Air 2	-0.57	-0.85	-0.23	1.41	4.01	1.01	-0.58	0.05	0.81	0.61

Air 3	-0.61	-0.96	-0.27	1.70	4.99	1.36	-0.87	0.03	0.85	0.86
Air 4	-0.58	-0.90	-0.26	1.55	4.41	1.14	-0.76	0.01	0.70	0.72
Helium 1	-0.33	-0.71	-0.35	0.42	0.44	-0.61	-0.94	-0.70	-3.67	-0.58
Helium 2	-0.32	-0.69	-0.34	0.38	0.30	-0.63	-0.88	-0.68	-3.54	-0.59
Helium 3	-0.32	-0.68	-0.33	0.39	0.27	-0.64	-0.86	-0.67	-3.49	-0.58
Helium 4	-0.32	-0.65	-0.31	0.34	0.06	-0.68	-0.76	-0.64	-3.27	-0.59
HQ 1	0.27	-0.70	-0.10	0.16	-1.59	4.17	-1.49	4.31	-0.46	-0.42
HQ 2	0.25	-0.32	-0.05	0.02	-0.85	2.14	-0.87	3.92	-1.54	0.19
HQ 3	0.29	-0.05	0.11	-0.13	-0.40	1.90	0.04	4.19	-1.90	0.39
Q 1	3.23	1.70	-4.11	-0.21	0.73	-0.65	0.52	0.52	-0.16	0.01
Q 2	3.49	1.77	-4.06	-0.16	0.66	-0.64	0.42	0.59	0.05	0.03
Q 3	3.36	1.55	-3.77	-0.14	0.46	-0.70	0.25	0.34	0.30	-0.08
Red 1	-0.50	0.06	-0.21	0.51	-0.46	1.08	3.86	-0.77	-0.53	1.00
Red 2	-0.53	0.05	-0.36	0.53	-0.68	1.84	1.85	-0.26	-0.33	1.44
Red 3	-0.54	0.14	-0.32	0.40	-0.60	1.26	3.72	-0.90	-0.59	0.97
QF1	-1.14	1.96	0.06	-0.21	-0.06	0.04	-0.20	-0.12	-0.34	1.08
QF2	-1.13	1.93	0.11	-0.18	-0.09	0.10	0.09	-0.10	-0.20	0.87
QF3	-1.15	1.97	0.10	-0.26	-0.10	0.07	-0.15	-0.20	-0.46	1.72
QF4	-1.13	1.99	0.07	-0.27	-0.10	0.08	-0.01	-0.11	-0.48	1.20
QF5	-1.09	1.78	0.18	0.10	-0.07	0.04	0.05	-0.14	0.03	0.93
QF6	-1.05	1.61	0.11	0.03	-0.25	-0.09	0.12	-0.41	-0.70	3.01
QF7	-0.98	1.49	0.06	-0.03	-0.05	0.06	0.19	-0.17	-0.48	1.43
QF8	-1.03	1.62	0.06	-0.10	-0.06	-0.03	0.06	-0.23	-0.52	1.69
QF9	-0.93	1.38	0.05	-0.04	-0.13	-0.04	-0.07	-0.26	-0.81	1.53
QF10	-0.98	1.47	0.02	-0.16	-0.09	0.09	0.25	-0.19	-0.69	1.68
QH 1	-0.07	-0.61	0.04	1.14	-0.60	-1.03	-0.82	0.53	0.44	0.13
QH 2	-0.07	-0.59	0.11	1.25	-0.68	-1.66	-0.63	0.77	0.38	0.28
QH 3	-0.10	-0.50	0.09	0.99	-0.29	-1.25	-0.37	0.51	0.23	0.19
QH 4	-0.10	-0.59	0.09	1.36	-0.57	-1.69	-0.55	0.92	0.85	0.44
QH 5	-0.12	-0.56	0.06	1.21	-0.70	-1.64	-0.71	0.81	0.50	0.75
QH 6	-0.06	-0.62	0.08	1.25	-0.86	-1.57	-0.86	0.73	0.29	0.34
QH 7	-0.10	-0.57	0.14	1.40	-0.65	-1.35	-0.70	0.72	0.87	0.40
QH 8	-0.12	-0.51	0.13	1.15	-0.29	-1.57	-0.27	0.71	0.43	0.29
QH 9	-0.10	-0.43	0.15	1.04	-0.29	-1.37	-0.29	0.48	0.19	0.45
QH 10	-0.10	-0.41	0.19	0.99	-0.20	-1.15	-0.24	0.33	0.19	0.32
97Q1	0.20	-0.57	0.39	-1.84	0.44	-0.66	0.04	0.14	0.04	0.33
97Q2	0.19	-0.57	0.39	-1.91	0.42	-0.66	0.04	0.09	-0.18	0.29
97Q3	0.18	-0.57	0.41	-2.01	0.45	-0.68	0.05	0.10	-0.20	0.37
97Q4	0.23	-0.58	0.42	-2.09	0.42	-0.70	0.04	0.10	-0.07	0.35
97Q5	0.19	-0.56	0.41	-1.92	0.43	-0.62	0.05	0.09	-0.06	0.42
97Q6	0.22	-0.60	0.45	-2.06	0.49	-0.62	0.09	0.15	0.19	0.24

97Q7	0.18	-0.55	0.41	-1.95	0.45	-0.66	0.06	0.08	-0.07	0.41
97Q8	0.20	-0.58	0.37	-1.87	0.44	-0.59	0.06	0.13	-0.06	0.18
97Q9	0.19	-0.56	0.37	-1.88	0.39	-0.59	0.01	0.01	-0.06	0.38
97Q10	0.23	-0.60	0.44	-2.06	0.49	-0.64	0.06	0.12	-0.01	0.28
97Q11	0.24	-0.63	0.44	-2.06	0.53	-0.79	0.06	0.27	-0.05	0.10
97Q12	0.17	-0.63	0.40	-1.92	0.51	-0.67	0.11	0.34	-0.06	-0.01
97Q13	0.17	-0.55	0.32	-1.95	0.40	-0.61	0.06	0.05	0.00	0.50
98Q1	-0.02	-0.53	-0.50	0.02	-0.82	1.02	-0.60	-1.07	0.68	0.07
98Q2	-0.01	-0.48	-0.45	-0.05	-0.71	0.83	-0.57	-0.97	0.95	0.17
98Q3	-0.01	-0.53	-0.48	0.01	-0.85	1.03	-0.69	-0.91	0.89	0.14
98Q4	0.02	-0.50	-0.53	0.12	-0.77	0.62	-0.68	-1.16	0.76	0.26
98Q5	-0.02	-0.52	-0.52	0.15	-0.78	0.89	-0.59	-1.02	0.80	0.16
98Q6	-0.03	-0.54	-0.60	-0.02	-0.80	0.82	-0.75	-1.09	0.45	0.05
98Q7	0.00	-0.56	-0.63	0.02	-0.72	0.65	-0.69	-1.12	0.25	-0.15
98Q8	-0.02	-0.58	-0.61	0.04	-0.76	0.80	-0.68	-1.15	0.39	-0.13
98Q9	-0.03	-0.56	-0.56	0.05	-0.74	0.72	-0.68	-1.23	0.36	0.08
98Q10	0.01	-0.59	-0.58	0.02	-0.67	0.74	-0.60	-1.19	0.46	-0.33
98Q11	-0.03	-0.61	-0.61	0.08	-0.74	0.70	-0.70	-1.13	0.42	-0.32
98Q12	-0.03	-0.58	-0.59	0.07	-0.62	0.55	-0.58	-1.19	0.33	-0.31
98Q13	-0.01	-0.56	-0.51	-0.05	-0.68	1.08	-0.45	-0.76	0.76	-0.38
98Q14	-0.03	-0.56	-0.52	0.13	-0.66	0.76	-0.48	-1.16	0.69	-0.26

Appendix V

Agglomeration schedule for PC scores extracted from correlation coefficients matrix all samples of HFC 134a analysed by GC ECD.

	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
Stage	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	77	79	0.002	0	0	6
2	41	42	0.004	0	0	18
3	3	4	0.012	0	0	57
4	76	82	0.021	0	0	19
5	74	75	0.031	0	0	19
6	77	81	0.043	1	0	8
7	95	96	0.057	0	0	17
8	77	85	0.073	6	0	30
9	33	35	0.091	0	0	12
10	73	80	0.11	0	0	30
11	92	93	0.13	0	0	29
12	32	33	0.16	0	9	54
13	19	20	0.18	0	0	51
14	91	94	0.21	0	0	29
15	86	90	0.24	0	0	20
16	83	84	0.27	0	0	39
17	95	97	0.31	7	0	41
18	40	41	0.35	0	2	48
19	74	76	0.39	5	4	33
20	86	88	0.43	15	0	31
21	53	56	0.47	0	0	53
22	12	14	0.51	0	0	47
23	8	9	0.56	0	0	42
24	27	28	0.60	0	0	62
25	65	71	0.65	0	0	26
26	65	72	0.71	25	0	50
27	60	62	0.76	0	0	35
28	64	68	0.82	0	0	49
29	91	92	0.87	14	11	52
30	73	77	0.93	10	8	33
31	86	87	0.99	20	0	45
32	50	52	1.05	0	0	79
33	73	74	1.12	30	19	55
34	17	18	1.18	0	0	51
35	59	60	1.25	0	27	43
36	23	25	1.32	0	0	46
37	47	48	1.39	0	0	56
38	54	57	1.48	0	0	65

39	78	83	1.57	0	16	55
40	66	69	1.67	0	0	63
41	95	99	1.76	17	0	52
42	7	8	1.86	0	23	58
43	59	61	1.98	35	0	70
44	37	39	2.11	0	0	66
45	86	89	2.24	31	0	60
46	21	23	2.37	0	36	74
47	12	13	2.50	22	0	61
48	40	43	2.65	18	0	75
49	64	67	2.80	28	0	59
50	65	70	2.99	26	0	64
51	17	19	3.20	34	13	67
52	91	95	3.42	29	41	69
53	53	55	3.66	21	0	65
54	32	34	3.91	12	0	81
55	73	78	4.17	33	39	82
56	47	49	4.45	37	0	96
57	3	5	4.75	3	0	82
58	6	7	5.06	0	42	78
59	30	64	5.37	0	49	63
60	86	98	5.76	45	0	67
61	12	15	6.21	47	0	88
62	27	29	6.65	24	0	73
63	30	66	7.10	59	40	72
64	63	65	7.57	0	50	72
65	53	54	8.06	53	38	70
66	37	38	8.61	44	0	76
67	17	86	9.19	51	60	71
68	45	46	9.92	0	0	83
69	16	91	10.98	0	52	71
70	53	59	12.19	65	43	77
71	16	17	13.62	69	67	92
72	30	63	15.11	63	64	85
73	26	27	16.73	0	62	85
74	21	22	18.46	46	0	80
75	11	40	20.40	0	48	90
76	36	37	23.05	0	66	95
77	53	58	25.74	70	0	89
78	6	10	28.45	58	0	89
79	50	51	31.69	32	0	86
80	21	24	34.93	74	0	96
81	31	32	38.41	0	54	86
82	3	73	42.29	57	55	87

83	44	45	48.39	0	68	91
84	1	2	54.80	0	0	87
85	26	30	61.74	73	72	94
86	31	50	73.73	81	79	88
87	1	3	87.28	84	82	92
88	12	31	117.82	61	86	90
89	6	53	173.80	78	77	97
90	11	12	241.63	75	88	91
91	11	44	326.44	90	83	93
92	1	16	414.88	87	71	93
93	1	11	506.37	92	91	94
94	1	26	599.26	93	85	95
95	1	36	692.47	94	76	97
96	21	47	787.33	80	56	98
97	1	6	882.77	95	89	98
98	1	21	980.00	97	96	0